

Db 90 CSSYPTSGGLYYWSAMLAPRWRAPLASWMTGWFNIVQWAVTASVDFSLAQLIQVIVLLS 149

QY 156 --AAYSHYTSQDSWHVFLIYEGVALVLLFNLFALKRNPWWHEIG-----FGLTIALFVI 208

Db 150 TGGRNGGKGSDFVIGIHGGLFIHALLNSLPISVLSFQALWLLGVLMILI 209

QY 209 SFIAILARSNPKAPNSQVWTANSNYTG--WSDGVCFILGLSTSCFMFICLDAAMHLAE 265

Db 210 PLV-----STERATTKFVFTNTDNLGLGITSYATIFVLGLLMSQYTTTCYDASAHMTEE 264

QY 266 CTDAARTVPKAVVSAILIIGFCTAFPTYTIAVLYGITDLSILS----SAGVIPFETMTQSL 321

Db 265 TVDADKNGRGIISAIGISILFGWGYILGISYAVTDIPSLSETNNSGGYALAEIFYLAF 324

QY 322 RSL--SFATVLSGGI--VMAFFALNAVQETASRLTWSFARDNGLVFSTHLERIHPRWQV 377

Db 325 KNREGSGTGIVCLGVVAVAFECGMSSVTSNSRMAYAFSRDGMMPSPMLHKNVSR-EV 383

QY 378 PVWSLFATW--GILATCGCI-FLGSSSTAFNALVNSAVVLQQLSELIPIALLLYQKRPKF 434

Db 384 PI---NAVLSALISECMALTSIGLSIVAFQAMVSIATIGLYIAYAIPILRVTLARN-TF 439

QY 435 LPSTRAFLVPRGIGFLVNLVAVVFTSVTTVFFSEFPLTVPTAASTMNYSIAIGVALALGV 494

Db 440 VPG--PFESLKG-YGMVVGWVAVLWVVTISVLSPLVPAYPITAETLNYTPVAVAGLVAIL 496

QY 495 LNWWVHARKHYQGP 508

Db 497 SYWLF SARHWTGP 510

RESULT 4

O60113 PRELIMINARY; PRT; 542 AA.

AC O60113;

DT 01-AUG-1998 (TrEMBLrel. 07, Created)

DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)

DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)

DE HYPOTHETICAL AMINO-ACID PERMEASE C15C4.04C.

GN SPBC15C4.04C.

OS Schizosaccharomyces pombe (Fission yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

OC Schizosaccharomycetales; Schizosaccharomycetaceae;

OC Schizosaccharomycetes.

OX NCBI_TaxID=4896;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=972;

RA Lyne M., Rajandream M.A., Barrell B.G., Xiang Z., Aves S.;

RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.

CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).

CC -!- SIMILARITY: BELONGS TO THE AMINO ACID PERMEASE FAMILY.

DR EMBL; AL023290; CAA18895.1; -.

DR InterPro; IPR002293; AA_rel_permease_1.

DR Pfam; PF00324; aa_permeases; 1.

DR PROSITE; PS00218; AMINO_ACID_PERMEASE; 1.

KW Hypothetical protein; Transport; Amino-acid transport; Transmembrane.

FT TRANSMEM 95 115 POTENTIAL.

FT TRANSMEM 217 237 POTENTIAL.

FT TRANSMEM 255 275 POTENTIAL.

FT TRANSMEM 298 318 POTENTIAL.

FT TRANSMEM 348 368 POTENTIAL.

FT TRANSMEM 402 422 POTENTIAL.

FT TRANSMEM 425 445 POTENTIAL.

FT TRANSMEM 469 489 POTENTIAL.

FT TRANSMEM 500 520 POTENTIAL.

FT DOMAIN 396 399 POLY-THR.

SQ SEQUENCE 542 AA; 59726 MW; 17D9B15C04299468 CRC64;

Query Match 16.5%; Score 445; DB 3; Length 542;

Best Local Similarity 29.3%; Pred. No. 3.8e-25;

Matches 146; Conservative 83; Mismatches 229; Indels 40; Gaps 16;

QY 36 PLEKKQFGTITIVSLAFVICNSWAGISLQALLAGGPVTLTYGILISTLVYICIAFSL 95

Db 55 PVKREFSTWATFSFAFSISGLFATVVTYSYPLISGGAPSVCWLIAGAGCMCIASLV 114

QY 96 AELTSVYPTAGGOYHFASILAPKSNRSISYVCGVLVSLLSWIAIGSSVTMIPAQQIPALI 155

Db 115 AELVSAYPTSGGLYFTCKDLVPARSPMPVAVVWGLNLGQAGVSSTDWCAQ---LLL 171

QY 156 AAYSHT---YSQDSWHVFLIYEGVALVLLFN-LFALKRNPWVHEI-GFGLT--IALFV 207

Db 172 AAVSISTDLKYIPTNQHIV---GVMAAIVVFHGLVNSLSTRDLDRITRFYATFHLIVLV 227

QY 208 ISFIAILARSNPKAPNSQVWTANSNYTGWSD-GVCFILGLSTSCFMFICLDAAMHLAEEC 266

Db 228 VCMICLLAKCPKFNFGKYVFTDVQASSGWHPIGFSFLFGFLSVAWCMTDYDATAHIAEEI 287

QY 267 TDAARTVPKAVVSAILIIGFCTAFPTYTIAVLYGI-TDLDSILSSAGYIP---FETMTQSL 321

Db 288 ENAAVRAPNAIALALSITYVLGWVFNIVLAFMTGMTDLSINSELGQPVQAIFYNVLGKK 347

QY 322 RSLSFATVLSGGIVMAFFALNAVQETASRLTWSFARDNGLVFSTHLERIHPRWQVPVWS 381

Db 348 GSAF-TILSF--IINFTGITAMQANA-RTIWAFSRDQALPFSRYWYKINKTTTTPV-- 401

QY 382 LFATWGILATC--GCIFLGSSTAFNALVNSAVVLQQLSFLIPIALLLYQKRPKFLPST 438

Db 402 -IAVWLNWVFCIALNLIGLSIEAIEAIFSVCAIALDWSYVIPIACKLIFFGKRLNKKPG- 459

QY 439 RAFVLPRGIGF--LVNVLAVVFTSVTVFFSFPLTVPTAASTMNYSIAIGVALALGV 495

Db 460 -----PWNLGWASHEFNAYAVCWTAFAVSVIFLMPTRVPTPQNMYAVVVLAVGLLPSLV 514

QY 496 NWVHARKHYQGP 513

Db 515 YWWSGARKSYIGPRINVD 532

RESULT 5

Q94CQ2 PRELIMINARY; PRT; 556 AA.

AC Q94CQ2;

DT 01-DEC-2001 (TrEMBLrel. 19, Created)

DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE P0660F12.28 PROTEIN.

GN P0660F12.28.

OS Oryza sativa (Rice).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

OC Ehrhartoideae; Oryzaceae; Oryza.

OX NCBI_TaxID=4530;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. NIPPONBARE;

RA Sasaki T., Matsumoto T., Yamamoto K.;

RT *Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC clone:P0660F12.28;

RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AP003687; BAB63862.1; -.

SQ SEQUENCE 556 AA; 59916 MW; 4B13A088F303A197 CRC64;

Query Match 16.1%; Score 432.5; DB 10; Length 556;

Best Local Similarity 28.2%; Pred. No. 3.3e-24;

Matches 142; Conservative 92; Mismatches 213; Indels 57; Gaps 16;

QY 39 KKQFGTITIVSLAFVICNSWAGISLQALLAGGPVTLTYGILISTLVYICIAFSLAEL 98

Db 76 KRGLSVSNFAFSFSIISVLTGVTGTTTGTGLRYGGPVSMTLGWLVAFAFNGCVALSMAEI 135

QY 99 TSVYPTAGGOYHFASILAPKSNRSISYVCGVLVSLLSWIAIGSSVTMIPAQQIPALI--- 155

```
Db 136 CSAYPTSGGLYYWSAKLAGKEWAPLASWVTGWFNIMGOWACTTSDVDFSLAQLLQVILLA 195
QY 156 --AAYSHTYSQDSWHVFLIYEGVALVLLFNLFALKRNPVWHEIGFGLTIALFVISFIAI 213
Db 196 TGGANGGGYMASKYVVLAIYSAILLHGLNSLPRLWLSWFGQLG-----AFWNV 245
QY 214 LARSNPKAPNSQVWTAWS--NYTGSQDGVCFIL--GLSTSCFMFIGLDAAMHLAEECTDA 269
Db 246 AAVAKERVSPEEFIFTHENAENGAGIHD-KAYILAAGLLMSQYSLIGYDTSAHIIETKNA 304
QY 270 ARTVPKAVVSAILIIGFCTAFPTYTIAVLYGITDLDLSILS----SAGYIPFETMTQSL--RS 323
Db 305 DWSGPIGIITSVLSTMEGWIYLIALTISIMTDIPYLLNPSNDAGGYAIAQAALYTSFHRRY 364
QY 324 LSFATVLSGGI--VMAFFALNAVQETASRLTWSFARDNGLVFSFTHLERIHPRWQVP--- 378
Db 365 GTGAGALACLGVIATAIFLCGSACITNSRMGYAFSRDAMPPLSRVWHRVDSR-EVPLNV 423
QY 379 VMSLFATWGLATCGCIFLGSSTAFNALVNSAVVLOQLSFLPIALLLYKRDPKFLPST 438
Db 424 VWLSVAVAFVMAITS--LGSQVAFQAMVSIAT----LGLLIACAL-----PVFERVT 469
QY 439 ---RAFVLPRG-----IGFLVNVLAUVVTSVTTVFFSFPLTVPTAASTMNYTSAIIGVA 489
Db 470 TARRSFV--RGPFHGLGKYGVIVGVVWVWVATVTVLFSLPVAYPVAKETNYTPVAVGGV 527
QY 490 LALGVLNVVWVHARKHYQGPHLELD 513
Db 528 LLLSLVAVWLHARFVFGQPVTVNVD 551

RESULT 6
Q94CQ3 PRELIMINARY; PRT; 515 AA.
AC Q94CQ3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE P0660F12.27 PROTEIN.
GN P0660F12.27.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT clone:P0660F12.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003687; BAB63861.1; -.
SQ SEQUENCE 515 AA; 55279 MW; 97BEF49648BA32A1 CRC64;
```

Query Match 15.2%; Score 410.5; DB 10; Length 515;
Best Local Similarity 25.8%; Pred. No. 1.3e-22;
Matches 133; Conservative 92; Mismatches 219; Indels 71; Gaps 14;

```
QY 39 KKQFGTITIVSLAFVICNSWAGISGSLQALLAGGPVTLLYGILISTLVYICIAFSLAEL 98
Db 31 KRGLSLVSNFAFSFISIISVLTGVTTYGTGLRYGGPVSMTLGWLVSFAFNCVALSMAEI 90
QY 99 TSVYPTAGGQYHFAFASILAPKSNRSISYVCGLVLSLWIAIGSSVTMIPAQQIPALI--- 155
Db 91 CSAYPTSGGLYYWSAKLAGKEWAPLASWVTGWFNIVQWACTTSDVDFSLAQLLQVILLA 150
QY 156 --AAYSHTYSQDSWHVFLIYEGVALVLLFNLFALKRNPVWHEIGFGLTIALFVISFIAI 213
Db 151 TGGANGGGYMASKYVVLAIYSVILHGLNSLPRLWLSWFGQLG----- 195
```

```
QY 214 LARSNPKAPNSQVWTAWSNYTG-WSDGVCVFILGLSTSCFMFIGLDAAMHLAEECTDAART 272
Db 196 -----AFWNVAENGMGIIHQKAYILAVGLLMSQYSVIGYDTSAHMIEETKNADWS 244
QY 273 VPKAVWSAILIIGFCTAFPTYTIAVLYGITDLDLSILS----SAGYIPFETMTQSL--RSLSF 326
Db 245 GPMGIITSVGLSTNFCMIYLIALTISIMTDIPYLLNPSNDAGGYAIAQAALYTSFHRRYGTG 304
QY 327 ATVLSGGI--VMAFFALNAVQETASRLTWSFARDNGLVFSFTHLERIHPRWQVP---VMS 381
Db 305 AGALACLGVIATAIFLCGSACITNSRMGYAFSRDAMPPLSRVWHRVDSR-EVPLNVVWL 363
QY 382 LFATWGLA-----TCG--CIF----LGSSPAFNALVNSAVVLOQLSF 410
Db 364 SVAVAFVMAITVNSKSIATAAAAVNFSSCDLVVFSWQSLGSQVAFQAMVSVTTGLGLYIAY 423
QY 419 LIPIALLLYKRDPKFLPSTRAFVLPGRIGFLVNVLAUVVTSVTTVFFSFPLTVPTAAST 478
Db 424 ALPVFFRVTAR-KSFVPG--QFHLGR-YGLMVGWMAVWVWATVTVLFSLPVAYPVAKET 479
QY 479 MNYTSAIIGVALGVLNVWVHARKHYQGPHLELD 513
Db 480 FNYTPVAVGGVLLSLVAVWFHARFVFGQPVTVNVD 514

RESULT 7
Q94CQ6 PRELIMINARY; PRT; 521 AA.
AC Q94CQ6;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE P0660F12.24 PROTEIN.
GN P0660F12.24.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT clone:P0660F12.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003687; BAB63858.1; -.
SQ SEQUENCE 521 AA; 55915 MW; 858AB5BBF5EDA2C8 CRC64;
```

Query Match 14.6%; Score 394; DB 10; Length 521;
Best Local Similarity 26.4%; Pred. No. 2.2e-21;
Matches 133; Conservative 96; Mismatches 230; Indels 44; Gaps 18;

```
QY 39 KKQFGTITIVSLAFVICNSWAGISGSLQALLAGGPVTLLYGILISTLVYICIAFSLAEL 98
Db 32 KRGLSALSNAFSPANISVMGVTTTNTGLRYGGPVSMTLGWLVSFAVFNCCVALSMAEI 91
QY 99 TSVYPTAGGQYHFAFASILAPKSNRSISYVCGLVSL-SWIAIGSSVTMIPAQQIPALI-- 155
Db 92 CSAYPTSGGLYYWSAKLAGKQWASLWVWTFWNVGVGWAAL-ASVDFSLAQLLQVILL 150
QY 156 ---AAYSHTYSQDSWHVFLIYEGVALVLLFNLFALKRNPVWHEIGFGLTI---ALFVI 208
Db 151 STGGNGGGYMAKYTVTLAIY--AFILHLGIINSLPIENLSLFGHVGAIWNAAGIFVL 207
QY 209 S-FIAILARSNPKAPNSQVWTAWSNYTGSQDGVCFI-LGLSTSCFMFIGLDAAMHLAEEC 266
Db 208 TILIPAVAKDRPNIEFVTHLNTENGMGTHDKAYILAVGLLMSQYSVIGYDTSARHVEET 267
QY 267 TDAARTVPKAVVSAILIIGFCTAFPTYTIAVLYGITDLDLSILS----SAGYIPFETMTQSLR 322
Db 268 KNADRSPIGIITSVLFATVFGWIYLIALTSVWTDIPYLLSPSNDAGGY-----AIAQALY 323
```


Db 369 VAAIWTASILSVLFVWG-----STLVSVACTSAITIVVSCVIFLFLSFTVPVIVGLMLAWG 424
QY 431 DPKFLPSTRAFLVPRGIGFLVNLVAVFTSVTTVFFSFPLTVPVTAASTMNTYSAIIGVA- 489
Db 425 TPKW-DKMGPNWGRGVFMLFAVLSIV--SMILIF-----VIGIQP 462
QY 490 -----LALGVNLVWVHARKHYQGPHELDGRVUGA 519
Db 463 PNDWALYITVGGFILTAVWFAFERNRFGQPPPL---GDIIAA 501

RESULT 12
Q92NI8
ID Q92NI8 PRELIMINARY; PRT; 515 AA.
AC Q92NI8;
DT 01-DEC-2001 (TremBLrel. 19, Created)
DT 01-DEC-2001 (TremBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE PUTATIVE AMINO-ACID PERMEASE PROTEIN.
GN SMC01597.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21368234; PubMed=11474104;
RA Galibert F., Finan T.M., Long S.R., Puehler A., Abola P., Ampe F.,
RA Barloy-Hubler F., Barnett M.J., Becker A., Bolstard P., Bothe G.,
RA Boutry M., Bowser L., Buhrmester J., Cadieu E., Capela D., Chain P.,
RA Cowie A., Davis R.W., Dreano S., Federspiel N.A., Fisher R.F.,
RA Gloux S., Godrie T., Goffeau A., Golding B., Gouzy J., Gurjal M.,
RA Hernandez-Lucas J., Hong A., Huizar L., Hyman R.W., Jones T., Kahn D.,
RA Kahn M.L., Kalman S., Keating D.H., Kiss E., Komp C., Lelaure V.,
RA Masuy D., Palm C., Peck M.C., Pohl T.M., Portetelie D., Purnelle B.,
RA Ramsperger U., Surzycki R., Thebault P., Vandenbol M.,
RA Vorhoeiter F.J., Weidner S., Wells D.H., Wong K., Yeh K.-C., Batut J.;
RT "The composite genome of the legume symbiont Sinorhizobium meliloti."
RL Science 293:668-672(2001).
DR EMBL; AL591790; CAC46793.1; --
KW Complete proteome.
SQ SEQUENCE 515 AA; 55915 MW; 956BD9C62060A524 CRC64;

Query Match 11.4%; Score 307.5; DB 16; Length 515;
Best Local Similarity 22.7%; Pred. No. 5.8e-15;
Matches 120; Conservative 99; Mismatches 230; Indels 79; Gaps 19;
QY 37 LEKK--QFGTITIVSLAF-VICNSWAGISGSLQALLAGGPVTLTYGILISTLVYICIAF 93
Db 25 LERRMSQFSNFAV---SFSIICILSGGINSLAQATSGAGG-AAIGGWPLGCFISFVFAV 80
QY 94 SLAELTSVYPTAGGQYHFASILAPKINSISY-----VGLVSLLSWIAIGSSVTMIP 147
Db 81 AMAQISSAYPTAGGLYHWG-SILG---NRFTGWLTAWFNLLGLVTVLGAINVGTYYFFMG 136
QY 148 AQQIPALIAAYSHYTSQDSWHVFLIYEGVALVLLFNLFALKRNPWWHIEIGFGLT----- 202
Db 137 SFGTF-----YFGLDFTMTTRIVFLAIITGAQALVNHMGIGLTKLTD 179
QY 203-----IALFVISFIILARSNPKAPNSQVWTAWNSYTGWSDG-----VCFILGL 246
Db 180 FSGVJIFVTAILLALVCLAAADTYEI--SRLFT-PSNYSGETGGNVWMPQTS GTWVFLGL 236
QY 247 STSCFMF IGLDAAHMLAECTDAARTVPKAVVSAIIIGFCTAFPTYIAVLYGITDLSIL 306
Db 237 LLPITYITGDASAHTSBETVKAHSVPRGMTSSVLWAAALFGYIMLC.SFVLMIPDMDEAA 296
QY 307 SSAGYIPFETMTQSLRSLSFATVLSGCGIYMAFEALNAVQETASRLTWSFARDNGLVFST 366
Db 297 KQGNVFFWAMDAQVNPJ-IKDVLYLAIFVSQWLGLATVTSVSRMIFAFSRDGLPASK 355

QY 367 HLERIHPRQVPPVWSLFATWGILAT-----CGCIFLCSSTAFNALVNSAVVLOQLSFLIP 421
Db 356 ALSKVSPTYRTPVAAIW-TGSILAVLFWGSSLSVIGDTPTVYTIIVSCTVIFLFEFAIP 414
QY 422 IALLLYQKRDPKFLPSTRAFLVPRGIGFLVNLVAVFTSVTTVFFSFPLTVPVTAASTMNY 481
Db 415 ITIGLFAWGTSKW-DKMGPNWNLGEGMKLFAILSIV-----AMVLIEVLGIQPPNDWALY 468
QY 482 TSAIIGVALALGVNLVWVHARKHYQGPHEL-----ELDGRVVGAEFQVG 524
Db 469 IT--VGFLIVTGII-WFAFENRRRFRGPPIGDEVARRQAEIAAAEKAVG 513

RESULT 13
Q9C0Z0
ID Q9C0Z0 PRELIMINARY; PRT; 543 AA.
AC Q9C0Z0;
DT 01-JUN-2001 (TremBLrel. 17, Created)
DT 01-JUN-2001 (TremBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE PUTATIVE AMINO ACID PERMEASE.
GN SPAPB24D3.02C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972H-;
RA Collins M., Harris D., Wood V., Rajandream M.A., Barrell B.G.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL590582; CAC36898.1; --
DR InterPro; IPR002293; AA_rel_permease_1.
DR InterPro; IPR002027; Amino_acid_permease.
DR Pfam; PF00324; aa_permeases; 1.
SQ SEQUENCE 543 AA; 59257 MW; 1B414DA8EAA9E762 CRC64;

Query Match 11.3%; Score 303; DB 3; Length 543;
Best Local Similarity 24.1%; Pred. No. 1.3e-14;
Matches 123; Conservative 109; Mismatches 227; Indels 52; Gaps 19;
QY 39 KKQFGTITIVSLAF---VICNSWAGISGSLQALLAGGPVTLTYGILISTLVYICIAFSL 95
Db 44 KREFSLLAVFGQSGSMGLCPSLYG-SMAFSMNCGAGG---MVWSWFGATCLLPATPAL 99
QY 96 AELTSVYPTAGGQYHFASILAPKINSISYVCGLVSLLSWIAIGSSVTMIPAQQIPALI 155
Db 100 SELASSMPTSGSLYFWTAYLSPPKYRAPLSWFLGYVLALAYSTGFASTIYAAAGLVQATA 159
QY 156 AAYSHYTSQDSWHVFLIYEGV-----ALVVL-----LFNLFALKRNPWWHIEIGFGLTIA 204
Db 160 SVANPSYAPTKEYEYGIYVALSFACSAIILVLP TKFLARFSENV-----VFQI---CTIL 211
QY 205 LFVISFIILARSNPKAPNSQVWTAWNSYTGWSD-GVCFILGLSTSCFMFIGLDAAMHLA 263
Db 212 IFIIS-LAASSTSETRNTGSIYIFGNFENYSGWTNMGWSFILCFTTPVWVLSGFESCATIV 270
QY 264 EECTDAARTVPKAVVSAIII---GECTAFPTYIAVLYGITDLSILSSAGYIPEETMTQ 319
Db 271 EEAKNASKAAPTAIISLTVSLFMGFCIMI--TIAGTMG-HDFSSILNTPYGEPSQVLY 327
QY 320 SLRSLSFATVLSGCGIYMAFEALNAVQETASRLTWSFARDNGLVFTSLHLRIHPRQVVP 379
Db 328 NNLGKRGAVGSAVLIILALCFNCALSCLASSREIFAFAFARDKGLPGSWIFRKLTP-CGIPL 386
QY 380 WSLFATWGILATCGCIFLGSSTAFNALVNSAVVLOQLSFLIP-IALLLYQKRD-P-KFLPS 437
Db 387 NAILLVNLYTIIVGULMLVNVTAISSIFNLAIIAFFISYSLPLVCRLLFNRLNPKGFY-- 444
QY 438 TRAFVLPRGIGFLVNLVAVFTSVTTVFFSFPLTVPVTAASTMNTYSAIIGVALGLVNLW 497

```
Db 445 CGKFSKP-----ISIVAVAWLWFMALMLLFPSPYQNPKNKVENWNAIVVLGFTVFVFCVGY 498
QY 498 VV---HARKHYQGPHELDGRV---VGAEFQ 522
Db 499 YLPKYGGKTFKGPVKTVDENVTGVTVDFQ 529

RESULT 14
Q9RKNO
ID Q9RKNO PRELIMINARY; PRT; 389 AA.
AC Q9RKNO;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE POSSIBLE AMINO ACID/METABOLITE PERMEASE (FRAGMENT).
GN SCD17.01.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Brown S.P., Harris D.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.*";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL118515; CAB56378.1; -.
DR InterPro; IPR002293; AA_rel_permease_1.
DR InterPro; IPR002027; Amino_acid_permease.
DR Pfam; PF00324; aa_permeases; 1.
FT NON_TER 1
SQ SEQUENCE 389 AA; 41731 MW; FCBE62374516AE3F CRC64;

Query Match 10.5%; Score 284; DB 2; Length 389;
Best Local Similarity 26.1%; Pred. No. 2.4e-13;
Matches 97; Conservative 59; Mismatches 150; Indels 66; Gaps 13;

QY 171 LIYEGVALVLLFNLFALK-----RNPVHVEIGFGLTIALFVIFIAILARSNPKAPN 223
Db 35 LLFAAILLHGLLNTFGVRIVGLLNSVSVVWHVLG-----VAVIVGALTTFAPD 82

QY 224 -----SQVWTANSNYTGWSDGVCFIL-GLSTSCFMFIGLDAAMHLAEECTDAARTVPKAV 277
Db 83 HHQASVFVGFERNNTGWGSGVYVVLIGLLMAQYTTFTGYDASAHMTEETHDASTAGPKGI 142

QY 278 VSAIIIGFCTAPPYTIAYLGITDLSILSS-AGYIPFETMTQSLRSL-----FAT 328
Db 143 VRSIWTSWIAGFVLLGLLGTFTFAIQSDGALTSPTGAPPAQILLDALGATAGKLLLVVIGA 202

QY 329 VLSCGGIVMAFFALNAVQETASRLTWSFARDNGLVFSTHLERIHPRQVVPVWSLFATWGI 388
Db 203 QLFCG---MASVTAN-----SRMIYAFSRDGPALPYSHIWSVNPRTPTV---AAVW-- 248

QY 389 LATGCCIFLG-----SSTAFNALVNSAVVLQQLSFLIPLALLLYQ----KRDPKFLPSTR 439
Db 249 LAALAALVIGLPLYNVTAYAAVTSIAVIGLYIAYVIPLELRVRKGAAFERGPWHLGRWS 308

QY 440 AFVLPRGIGFLVNVLAUVFTSVTTVFESFPLTVPTAASIMNYTSAIIGVALALGVLNWV 499
```

```
Db 309 Q-----LVGVAVTWGVITVLFMLPQVSPVTWETFNYPAPIAVLAVLGFAATWMLV 359
QY 500 HARKHYQGPHE 511
Db 360 SARHWFLNPDHE 371

RESULT 15
Q9L1J2
ID Q9L1J2 PRELIMINARY; PRT; 476 AA.
AC Q9L1J2;
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)
DE PUTATIVE AMINO ACID TRANSPORTER.
GN SCC46.09C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Brown S.P., Harris D.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.*";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL139164; CAB75313.1; -.
DR InterPro; IPR002293; AA_rel_permease_1.
SQ SEQUENCE 476 AA; 49188 MW; B3D85592BC3FC0C3 CRC64;

Query Match 9.3%; Score 250.5; DB 2; Length 476;
Best Local Similarity 23.5%; Pred. No. 9.1e-11;
Matches 125; Conservative 85; Mismatches 209; Indels 113; Gaps 22;

QY 17 TTKNTEPAAAGGASESLNVPLE---KKQFGT-----ITIVSLAFVICSWAGIS 62
Db 4 TLRPAETAAPVPPDTPAPQAPSESRTLKRSIGVVGVTLLTLCVTPASTLFFVV---PDLF 60

QY 63 GSLQLALLAGGPVTLTYGILISTLVYICIAFSLAELTSVYPTAGQYHFASILAPKSINR 122
Db 61 GSL-----GTATAL-TIAGSLLCVAVAFCYSELCTLVPSAGGEYAMVSTMA---GR 108

QY 123 SISVYVCGILVSLLSWIAIGSSVTMPAQQIPALIAAYSHYTSQDSWHV--FLIYEGVALV 180
Db 109 LAGWLVFVLSLL-----VVM-----VPPVIAMCTADYLEPVLHLDPSLAGAGVMLLA 156

QY 181 LLFNLFAKRNPNVHVEIGFGLTIALFVIFIAILARSN-PKAPNSQVWTANSNYTGWSDG 239
Db 157 TLAGLLDLRANAWI--TGVFLVLEVIAAAVAVLGFHAERGPGLSVSMEVGGTGGADP 214

QY 240 VCFIL---GLSTSCFMFIGLDAAMHLAEECTDAARTVPKAVVSAIIIGFCTAPPYTI 296
Db 215 VTAMLVYSGLAIALFITQGFSTAVYLSELEENPRRNARTVLTATLAISTVILLVPVAIT 274

QY 297 YGITDLSI---LSSAGYIPFETMTQSLRSLSFATVLSGGIVMAFFALNAVQETASRL 352
Db 275 FGASDLAELTGGDISAMVTANSNSAVGTFVSLCVALAIVNAGIVM-----VIQN--SRV 326
```

QY	353	TWSEFARDNGLVFSTHLERIHPRWQVPV-----WSLFATWGILATGCGIFLGS	399
Dd	327	LEASARDKA-----WPAPVNGLLGRLGREGSPWVATLAVGPGAALC-YVDL	372
QY	400	STAFNALVNSAVVLQQLSFLIPTALLLYQKDPKFLEPSTRAEVLPRGIGELVNVLAVFT	459
Dd	373	DTLYGV--TGVSVTAMYLVLVAVALLSRRGAHRHHAWRMPLPVPPVLLIAYLVLT	429
QY	460	SVTTVFSEPLTVPTAASTMNTYSAILGVALALGVNLVVVHARKHYQGPHLE	511
Dd	430	Q-----QEASHLLTGSIAAVA---TLVWALYLR-----PRLD	459

Search completed: April 27, 2002, 07:57:47
Job time: 703 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2002, 07:59:39 ; Search time 35.22 Seconds
(without alignments)
577.165 Million cell updates/sec

Title: US-09-882-694A-8
Perfect score: 2692
Sequence: 1 MDSRPSGYGKGTQTTKN.....QGPHELDGRVVGAEFQVGP 525

Scoring table: BLOSUM62
Gapop 10.0 , Capext 0.5

Searched: 105224 seqs, 38719550 residues
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	681.5	25.3	563	1 HNM1_YEAST	P19807 saccharomyc
2	540.5	20.1	561	1 BIO5_YEAST	P53744 saccharomyc
3	454	16.9	567	1 YF21_SCHPO	Q9us40 schizosacch
4	429	15.9	557	1 YCQ4_SCHPO	O74537 schizosacch
5	420	15.6	544	1 YC9D_SCHPO	Q09887 schizosacch
6	409.5	15.2	551	1 AAP2_NEUCR	O59942 neurospora
7	383.5	14.2	571	1 UGA4_YEAST	P32837 saccharomyc
8	351.5	13.1	550	1 YAO8_SCHPO	Q10087 schizosacch
9	334	12.4	553	1 GPT1_CANAL	O74248 candida alb
10	321	11.9	618	1 YKR4_YEAST	P36029 saccharomyc
11	311.5	11.6	554	1 YCT3_SCHPO	O59813 schizosacch
12	274	10.2	591	1 YFYA_SCHPO	Q9ut18 schizosacch
13	225.5	8.4	461	1 YCJJ_ECOLI	P76037 escherichia
14	215.5	8.0	452	1 YEEF_ECOLI	P33016 escherichia
15	199.5	7.4	440	1 YJ99_MYCTU	Q10858 mycobacteri
16	197.5	7.3	596	1 AGP2_YEAST	P38090 saccharomyc
17	188.5	7.0	608	1 DIP5_YEAST	P53388 saccharomyc
18	181	6.7	636	1 CTR4_HUMAN	O43246 homo sapien
19	177.5	6.5	602	1 GAP1_YEAST	P19145 saccharomyc
20	177.5	6.5	657	1 CTR2_MOUSE	P18581 mus musculu
21	177	6.5	499	1 ANSP_ECOLI	P77610 escherichia
22	175	6.5	558	1 AGP3_YEAST	P43548 saccharomyc
23	174.5	6.5	465	1 Y093_RHIME	O87394 rhizobium m
24	174.5	6.5	658	1 CTR2_HUMAN	P52569 homo sapien
25	170.5	6.3	418	1 YJEH_ECOLI	P39277 escherichia
26	170	6.3	469	1 GABP_BACSU	P46349 bacillus su
27	169.5	6.3	574	1 YHC6_SCHPO	O60170 schizosacch
28	169	6.3	497	1 ANSP_SALTY	P40812 salmonella
29	168.5	6.3	603	1 HIPI_YEAST	P06775 saccharomyc
30	168	6.2	481	1 YJ79_MYCTU	Q10875 mycobacteri
31	167.5	6.2	445	1 YHEM_ECOLI	P45539 escherichia
32	165	6.1	624	1 CTR1_RAT	P30823 rattus norv
33	164.5	6.1	488	1 LYSP_ECOLI	P25737 escherichia

34	164	6.1	622	1 CTRL_MOUSE	Q09143 mus musculu
35	162	6.0	466	1 GABP_ECOLI	P25527 escherichia
36	162	6.0	571	1 CAN1_CANAL	P43059 candida alb
37	161.5	6.0	458	1 YDGF_BACSU	P96704 bacillus su
38	159.5	5.9	611	1 LYPI_YEAST	P32487 saccharomyc
39	159	5.9	458	1 PHEP_ECOLI	P24207 escherichia
40	159	5.9	573	1 INAI_TRIHA	P34054 trichoderma
41	157.5	5.9	533	1 LAT2_RAT	Q9wvr6 rattus norv
42	156	5.8	445	1 YUDE_ECOLI	P39269 escherichia
43	156	5.8	531	1 LAT2_MOUSE	Q9qwx9 mus musculu
44	154.5	5.7	457	1 AROP_SALTY	Q9l4i0 salmonella
45	152.5	5.7	470	1 CYCA_ECOLI	P39312 escherichia

ALIGNMENTS

RESULT 1
HNM1_YEAST
ID HNM1_YEAST STANDARD; PRT; 563 AA.
AC P19807;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Choline transport protein.
GN HNM1 OR CTR1 OR CTR OR YGL077C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=X2180-1B;
RX MEDLINE=90368823; PubMed=2203793;
RA Nikawa J.-I., Hosaka K., Tsukagoshi Y., Yamashita S.;
RT "Primary structure of the yeast choline transport gene and regulation
of its expression.";
RL J. Biol. Chem. 265:15996-16003(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=97435481; PubMed=9290212;
RA Rieger M., Brueckner M., Schaefer M., Mueller-Auer S.;
RT "Sequence analysis of 203 kilobases from Saccharomyces cerevisiae
chromosome VII.";
RL Yeast 13:1077-1090(1997).
CC -!- FUNCTION: SOLE CHOLINE TRANSPORTER IN YEAST.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- MISCELLANEOUS: INHIBITION OF ACTIVITY BY INTRACELLULAR CHOLINE.
CC -!- SIMILARITY: BELONGS TO THE AMINO ACID PERMEASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
or send an email to license@isb-sib.ch).

CC EMBL; J05603; AAA34537.1; -;
CC EMBL; 272599; CAA96782.1; -;
CC PIR; S11175; S11175.
CC SGD; S0003045; HNM1.
CC InterPro; IPR002293; AA_rel_permease_1.
CC InterPro; IPR002027; Amino_acid_permease.
CC Pfam; PF00324; aa_permeases; 1.
CC PROSITE; PS00218; AMINO_ACID_PERMEASE_1; 1.
KW Transport; Amino-acid transport; Transmembrane.
FT TRANSMEM 58 78 POTENTIAL.
FT TRANSMEM 88 108 POTENTIAL.
FT TRANSMEM 183 203 POTENTIAL.
FT TRANSMEM 205 226 POTENTIAL.
FT TRANSMEM 256 276 POTENTIAL.

```
FT TRANSMEM 294 314 POTENTIAL.
FT TRANSMEM 343 363 POTENTIAL.
FT TRANSMEM 399 417 POTENTIAL.
FT TRANSMEM 427 445 POTENTIAL.
FT TRANSMEM 466 486 POTENTIAL.
FT TRANSMEM 492 512 POTENTIAL.
FT CARBOHYD 7 7 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 20 20 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 248 248 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 341 341 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 563 AA; 62056 MW; 1A71E1D146DB7807 CRC64;

Query Match 25.3%; Score 681.5; DB 1; Length 563;
Best Local Similarity 33.2%; Pred. NO. 9.7e-37;
Matches 156; Conservative 95; Mismatches 210; Indels 9; Gaps 6;

QY 39 KQFGTITIVSLAFVICSWAGISGSLQALLAGGPVTLTYGILISTLVYICIAFSLAEL 98
Db 55 RKFSLSILGVGFLTNSWFGISTSMVAGISSGPMIVYGIIVIALISICITSLGEL 114

QY 99 TSVYPTAGQVHPASILAPKINSRISYVCGVLSLWSIAIGSSVTMIPAAQIPALIAAY 158
Db 115 SSAYPHAGGQVWSLKLAPPKYKRAAYMCGSFAYAGSVFTSASTTSLVATEVVGMYALT 174

QY 159 SHYTSQDSWHVFLIEGVALVLLFNLFALKRNPVWVEIGFLTIALFVIFIAILARSN 218
Db 175 HPEFIPKRWHPVCFELHLFLMFFNCYG-KSLPISSSSLYISLSSFTTITVTLACSH 233

QY 219 PKAPNSQ-VMTAWSNYTGWSD-GVCFILGLSTSCFMFIGLDAAMHLAEECTDAARTVPKA 276
Db 234 GKFNDAKVFATFNNETGWKNGGIAFIVGLINPAWSPFCLDCATHMAFEVEKPERVPIA 293

QY 277 VVSAIIIGECTAFPTYTIAVLYGIDTLDLSILSSAGYIP-FETMTQSLRSLSPATVLSGCGI 335
Db 294 IMGTVAIGFVTSFCYVIAMFFSIQDLDAVLSSTTGAPILDIYNQALGNKSGAIFLGLIL 353

QY 336 VMAFFALNAVQETASRLTWSFARDNGLVFSSTHLERIHPRWQVVPVWSLFPATWGILATCGCI 395
Db 354 FTSGFCVIACTHQAARLCSWFARDNGLPLSRWSQVNPHTGVPLNAHLMSCAWITLIGLL 413

QY 396 FLGSSAFNALVNSAVVLOQLSFLIPIALLLYQKRPKFLPSTRAFLPRGIGFLVNVLA 455
Db 414 YLASSTAFQSLITGCI AFLLSYIIPVICLLAKKRNIAHGP----FWLGK-FGFESNIVL 468

QY 456 VVFTSVTVFFSFPLTVPTAASVMNYSALIGVALALGVNLVWVHARKHY 505
Db 469 LGWTVFSWVFFSFPPVLPVTKDNMNVVGVVIGYTAYSILYKYGKKEF 518

RESULT 2
BIOS_YEAST
ID BIOS_YEAST STANDARD; PRT; 561 AA.
AC P53744;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Transmembrane regulator of KAPA/DAPA transport.
GN BIOS OR YNR056C OR N3502.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-FL100;
RA Phalip V., Jeltsch J.M., Lemoine Y.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Duesterhoeft A., Floeth M., Fritz C., Heuss-Neitzel D.,
RA Hilbert H., Moestl D.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
```

```
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: SOME, TO HNMI/CTR1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U90440; AAB50012.1; -.
DR EMBL; Z71671; CAA96337.1; -.
DR SGD; S0005339; BIO5.
DR InterPro; IPR002293; AA_rel_permease_1.
KW Transmembrane.
FT TRANSMEM 50 70 POTENTIAL.
FT TRANSMEM 78 98 POTENTIAL.
FT TRANSMEM 161 181 POTENTIAL.
FT TRANSMEM 205 225 POTENTIAL.
FT TRANSMEM 231 251 POTENTIAL.
FT TRANSMEM 282 302 POTENTIAL.
FT TRANSMEM 322 342 POTENTIAL.
FT TRANSMEM 370 390 POTENTIAL.
FT TRANSMEM 440 460 POTENTIAL.
FT TRANSMEM 462 482 POTENTIAL.
FT TRANSMEM 508 528 POTENTIAL.
FT TRANSMEM 541 561 POTENTIAL.
SQ SEQUENCE 561 AA; 62428 MW; 7FA5A64FA0D4250F CRC64;

Query Match 20.1%; Score 540.5; DB 1; Length 561;
Best Local Similarity 25.9%; Pred. NO. 1e-27;
Matches 136; Conservative 120; Mismatches 202; Indels 67; Gaps 14;

QY 29 ASESINVPLEKKQFGTITIVSLAFVICSWAGISGSLQALLAGGPVTLTYGILISTLVY 88
Db 33 SSERSEVKFD-KHFNWWSLLGIAFSLSCSVWGISASMAVGASGGPLLIYGLIIAFFS 91

QY 89 ICIAFSLAELTSVYP-TAGQYHFASILAPKSI----- 120
Db 92 LMCGISLGDFAAILPNSSGGSFVWLKMLEQESVTLKTEYEDPSDDDEEVLENYCOTFN 151

QY 121 -----NRSISYVCGVLSLWSIAIGSSVTMIPAAQIPALIAAYSHTYSQDSWHVFLIY 173
Db 152 VEVSSKFKQVSSMVVGLNYPFAIFTTASICSSLSMSCIGIHKLLHPDYELKHWVFGY 211

QY 174 EGVALVLLFNLFALKRNPWVHETGFLTIALFVIF-IAILARSN-----PKAPNSQV 226
Db 212 ECINAVLTLFNIYSTPL-PYISQFGLYTSLSFAMTFIICIVSRSDNTVDPWPKASN--I 268

QY 227 WTAWSNYTGW-SDGVCFILGLSTSCFMFIGLDAAMHLAEEC--TDAARTVPKAVVSAIII 283
Db 269 FGSFDNQTCWNSSGMAFVGLVNPVIAFVIGIDSATHMIDEVGYSKSRFLVPKVIITIIV 328

QY 284 GFCTAFPTYTIAVLYGITDLSILSSAGYIPF-ETMTQSLRSLSPATVLSGCGIVMAFFAL 342
Db 329 GFVTSFIYCVGLFFCITDOTAVVES--ILPIVEIFYQATGNRNLVSFLOCMCITTFGVSG 386

QY 343 NAVQETASRLTWSFARD-----NGLVFSSTHLER---IHPRWQVVPVWSLFPATWGILATCG 393
Db 387 IASGTWQSRILQSFCKSYAPFYKESGLGNKSLKLLAVLTGPKSPLYAHFLSQICVTIIG 446

QY 394 CIFLGSSTAFNALVNSAVVLOQLSFLIPIALLLYQKRPKFLPSTRAFY-----LPRGI 447
Db 447 CIFMGSSSTAFNAITACITLLLMYSYAVPSFIFLVFKKEKFIHRIESDNCVSRPNRRM 506

QY 448 GFLVNVLAHVTSVTTVFFSFPLTVPTAASTMNTVTSAILGVALAL 492
Db 507 SMIPHIICILWTLFCLVFLSFPTYTLPTAGNMNTVSVVYAVVFCI 551

RESULT 3
```



```
Db 270 SDWNGWQFCLAGFMPAVWTIGSPDSCVHQSEAKDAKKSVPIGIISIAVCWILGWLI 329
QY 293 IAVLYGIT-DLDSILSSA-GYIPFETMTQSLR---SLSFATVLSGGIVMAFFALNAVQE 347
Db 330 ICLMACINPDIDSVLDSKYGFALAQIYDLSLKKWIAFMSL-----IAFCQFLMGASIT 384
QY 348 TA-SRLTWSFARDNGLVFSFTHLERIHPRWQVPVWSLFWATWGILATC-GCIFLG-----SS 400
Db 385 TAVSRQWAFSRDNGPLSKYIKRVDKSYVPPFA-----ILAACVGSLLGLCLIDD 438
QY 401 TAFNALVNSAVVLQQLSEFLIPTALLLYQKRPKFLPSTRAFLVPRGIGFLVNVLAUVFTS 460
Db 439 AATDALFSLAVAGNNLAWSTPTVRLTSGRD-LFRPG--PFYLGKIWSPIVANTGVAFQL 495
QY 461 VTTVFFSFP-----LTVPTAASTMNYSALIGVAL-ALGVLNVVHARKHYQGPHLELDG 514
Db 496 FIILVMFPSSQHQGIT----KSTMNY-ACVIGPGIWILAGIYYKYKKKYYHGPATNLS 550
QY 515 ----RVVGAE 520
Db 551 DDYTEAVGAD 560

RESULT 8
YAO8_SCHPO
ID YAO8_SCHPO STANDARD; PRT; 550 AA.
AC Q10087;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative amino-acid permease Clld3.08c.
GN SPAC11D3.08C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Connor R., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE AMINO ACID PERMEASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z68166; CAA92309.1; -.
DR InterPro; IPR002293; AA_rel_permease_1.
DR InterPro; IPR002027; Amino_acid_permease.
DR Pfam; PF00324; aa_permeases; 1.
DR PROSITE; PS00218; AMINO_ACID_PERMEASE_1; 1.
KW Hypothetical protein; Transport; Amino-acid transport; Transmembrane.
FT TRANSMEM 53 73 POTENTIAL.
FT TRANSMEM 78 98 POTENTIAL.
FT TRANSMEM 127 147 POTENTIAL.
FT TRANSMEM 177 197 POTENTIAL.
FT TRANSMEM 201 221 POTENTIAL.
FT TRANSMEM 246 266 POTENTIAL.
FT TRANSMEM 283 303 POTENTIAL.
FT TRANSMEM 336 356 POTENTIAL.
FT TRANSMEM 386 406 POTENTIAL.
FT TRANSMEM 411 431 POTENTIAL.
FT TRANSMEM 453 473 POTENTIAL.
FT TRANSMEM 481 501 POTENTIAL.
SQ SEQUENCE 550 AA; 59918 MW; 3DD1EF6003896E4C CRC64;
```

```
Query Match 13.1%; Score 351.5; DB 1; Length 550;
Best Local Similarity 23.4%; Pred. No. 1.2e-15;
Matches 129; Conservative 107; Mismatches 238; Indels 77; Gaps 16;

QY 10 EKGGTROTTKNTETAAAGGASESLN-----VPLEKKQFGTTITIVSLAFVICNSW 58
Db 5 DKESLKQKQKQWENEVELGESREANQDDELLMSLGKYPEFTREFSYVSIFGQSPGSMGLC 64
QY 59 AGISGSLQALLAGGPVTLTYGILISTLVYICIAFLAELTSVYTAGGQVHFASILAPK 118
Db 65 PAMAGSLIFSMNCGG-GMVWSWIGICICLIPVSIISLGELASSMPTSGGLYFWIFTLASP 123
QY 119 SINRSISYVCGVLSLWSWIAIGSSVTMIPAAQIIPALIAAYSHTYSQDSWHVFLIYEGVAL 178
Db 124 SSRAFLCWVCGYVSVLGYATYASTVYSASSMVQALAVIGSPSYPTKYEQYGIYAALLF 183
QY 179 VLLFNLFALKRNPWVHEIGFGLTIALFVISFIALIARSNPKAPNSQ--VWTANSNYTGW 236
Db 184 VISAMTAIPSRVIAKVNIINTFQFLVSIILLIALAAGSDSTTRNSGSIQDFTNYSGW 243
QY 237 SD-GVCFILGLSTSCFMFGLDAAMHLAEECTDAARTVPKAVVS----AIIIGFCTAPPY 291
Db 244 SNMGWAFILSFTTPVWVVSFGFESSAAVAEESTNAAKAANPFAMISSLGVATILGWICIVI-- 301
QY 292 TIAVLYGITDLSIL-SSAGYIPFETMTQSLRS-----LSFATVLSGGIVMAFFA 341
Db 302 TVVATMG-HDFNAILGSSSLGQPVAVQLVNVNNGKALGIFSLVIALCLNCLISLLIA--- 357
QY 342 LNAVQETASRLTWSFARDNGLVFSFTHLERIHPRWQVPVWSLFWATWGILATCGCIFLGSST 401
Db 358 -----ASREVFAFCRDGGIPGSRYL-RLLTQKQVPLNAILLVLLYSLLVGLLILVNV 409
QY 402 AFNALVNSAVVLQQLSEFLIPI-ALLLYQKRPKFLPSTRAFLVPRGIGFL-----VNV 454
Db 410 AISSVFNLAIIALYIAYSGPLMCRFVYKFKQP-----GVFYCKWSKPAALW 456
QY 455 AVVFTSVTTVFFSFPPLTVPTAASTMNYSALIGVALGVLNVVHARKHYQGPHLELDG 514
Db 457 SLVMMWFMI LMLLPQYQKPNQDEMNAIVLGVFMVFCVV-----YYLP----- 502
QY 515 RVVGAEFQVGP 525
Db 503 KIGGKTFTTGP 513

RESULT 9
GPT1_CANAL
ID GPT1_CANAL STANDARD; PRT; 553 AA.
AC O74248;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Putative polyamine transporter.
GN GPT1.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
RN [1]
RP SEQUENCE FROM N.A.
RA McNemar M.D., Gorman J.A., Buckley H.R.;
RT "Isolation and sequence of the GPT1 gene of Candida albicans, encoding
RT a putative polyamine transporter.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -1- SIMILARITY: BELONGS TO THE AMINO ACID PERMEASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; AF080132; AAC31569.1; -.
DR InterPro; IPR002293; AA_rel_permease_1.
DR InterPro; IPR002027; Amino_acid_permease.
DR Pfam; PF00324; aa_permeases; 1.
DR PROSITE; PS00218; AMINO_ACID_PERMEASE_1; FALSE_NEG.
KW Transport; Transmembrane.
FT TRANSMEM 43 63 POTENTIAL.
FT TRANSMEM 68 88 POTENTIAL.
FT TRANSMEM 133 153 POTENTIAL.
FT TRANSMEM 163 183 POTENTIAL.
FT TRANSMEM 190 210 POTENTIAL.
FT TRANSMEM 240 260 POTENTIAL.
FT TRANSMEM 280 300 POTENTIAL.
FT TRANSMEM 333 353 POTENTIAL.
FT TRANSMEM 397 417 POTENTIAL.
FT TRANSMEM 453 473 POTENTIAL.
FT TRANSMEM 485 505 POTENTIAL.
SQ SEQUENCE 553 AA; 60633 MW; 197DACB76C5FBC9E CRC64;

Query Match 12.4%; Score 334; DB 1; Length 553;
Best Local Similarity 23.5%; Pred. No. 1.6e-14;
Matches 120; Conservative 91; Mismatches 233; Indels 56; Gaps 18;

QY 39 KKQFGTITIVSLAFVICNSWAGISGLQALLAGGPVTLGYLILSTLVYICIAFSLAEL 98
Db 37 RREFSLWSIFAVSPSVLGLLPSSIAACFDYQQLVVGMSPLPW--LIAMIFITSVAYSMAEI 94
QY 99 TSVYPTAGGQYHFASILAPKSIINRSISYVCGVLSLWSIAIGSSVTMIPAAQIPALIAAY 158
Db 95 ASAFPCSAAGTPYAVSQAAPK-----KYASFLTWTCTWNWSCQITAAPS-----VSY 141
QY 159 S-----HTYSQDSW-----HVFLIYEGVALVLLFNLFLALKRNPWV---HEIGFGLT 202
Db 142 SCACMMLALHSFTDPSPVASNAQIFGLTTGIVLCAPMACFPFK---WVARFSSAGTTCN 198
QY 203 IALFVISFIAILA---RSNPKA-----PNSQVWTAWSNYTGWSDGVCFILGLSTSCFMF 253
Db 199 IVFLVVVFIMILGGNKRDQIKEGISKFNSNSTAW-GLDNQAEWPTGLSFLISFMGVWAM 257
QY 254 IGLDAAMHLAEECTDAARTVPKAVVSAIIIGFCTAFPTYIAVLYGITDLDLSILS---SAG 310
Db 258 SGYDSPFHAEKCSNAVAAPRAIVLTSTVGGGLIGFMFMAIAYTLVDLNLQISADPESLG 317
QY 311 --YIPFETMTQSLRSLSPATVLSGCGGIVMAFFNALNAVQETASRLTWSFARDNGLVFSTHL 368
Db 318 QPFVYTLTQIMDKNLNVIGAFALT---IISFFMAQNCLLASSRVTYAYARDGLFPLSGIW 374
QY 369 ERIHPRWQVPVWSLFATWGLATCGCIFLGSSTAFNALVNSAVLQQLSFLIPIALLLYQ 428
Db 375 KKVSPKQTPTPINAVIMNFIVEBELLLLIFGGDVSIGSFISGALAGFISFTMPTLLKITY 434
QY 429 KRDPKFLPSTRAFVLPRGIGFL---VNVLAVVFTSVTTVFPSPPLTVPTAAST---MNYT 482
Db 435 AR-KTFQPG-----PWNLGKWEPIGWVSFAVGLMVPILCFP-TVKGADLTPTTEMNWT 486
QY 483 SAIGVALALGVLNWNVHARKHYQGPHLEL 512
Db 487 CLVYFGLILLTTIWFVYVYARRVYGPRTNI 516

RESULT 10
YKR4_YEAST
ID YKR4_YEAST STANDARD; PRT; 618 AA.
AC P36029;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical amino-acid permease in STE3-GIN10 intergenic region.

GN YKL174C OR YKL639.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID-4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C;
RX MEDLINE-94378719; PubMed-8091858;
RA Vandenbol M., Bolle P.-A., Dion C., Portetelle D., Hilger F.;
RT "Sequencing and analysis of a 20.5 kb DNA segment located on the left
RT arm of yeast chromosome XI.";
RL Yeast 10:S25-S33(1994).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -!- SIMILARITY: BELONGS TO THE AMINO ACID PERMEASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; Z26878; CAA81512.1; -.
DR EMBL; Z28174; CAA82016.1; -.
DR PIR; S38004; S38004.
DR PIR; S38406; S38406.
DR PIR; S44586; S44586.
DR SGD; S0001657; YKL174C.
DR InterPro; IPR002293; AA_rel_permease_1.
DR InterPro; IPR002027; Amino_acid_permease.
DR Pfam; PF00324; aa_permeases; 1.
DR PROSITE; PS00218; AMINO_ACID_PERMEASE_1; 1.
KW Hypothetical protein; Transport; Amino-acid transport; Transmembrane.
FT TRANSMEM 61 84 POTENTIAL.
FT TRANSMEM 91 110 POTENTIAL.
FT TRANSMEM 132 148 POTENTIAL.
FT TRANSMEM 155 171 POTENTIAL.
FT TRANSMEM 180 200 POTENTIAL.
FT TRANSMEM 212 231 POTENTIAL.
FT TRANSMEM 298 317 POTENTIAL.
FT TRANSMEM 343 367 POTENTIAL.
FT TRANSMEM 403 419 POTENTIAL.
FT TRANSMEM 426 449 POTENTIAL.
FT TRANSMEM 465 486 POTENTIAL.
FT TRANSMEM 499 516 POTENTIAL.
SQ SEQUENCE 618 AA; 69242 MW; 035CB927C624CE74 CRC64;

Query Match 11.9%; Score 321; DB 1; Length 618;
Best Local Similarity 24.6%; Pred. No. 1.2e-13;
Matches 116; Conservative 99; Mismatches 208; Indels 48; Gaps 15;

QY 37 LEKKQFGTITIVSLAFVICNSWAGISGLQALLAGGPVTLGYLILSTLVYICIAF--- 93
Db 51 LDKSLLSRGSIVGLGLGLMSPVLGMCTSMAGLINGGPLTLMGLFLISG---VCIWFSSSL 107
QY 94 SLAELTSVYPTAGGQYHFAS-ILAPKSIINRSISYVCGVLSLWSIAIGSSVTMIPAAQIP 152
Db 108 SLGEIVSKFPM---ELHVGSAMLAPEKCLKVCSWYTGWLMLIGNWTMSTSIFFAGAQLTI 164
QY 153 ALIAAYSHTYSQDSWHVF-----LIYEGVALVLLFNLFLALKRNP-----WVHEIGFGL 201
Db 165 SLILMTNSNLISEAHLIFTYIVFVLYVTVVGLVNLKFAFRTIETINKVCVYWI----- 217
QY 202 TIALFVISFIAILA---RSNPKAPNSQVWTAWSNYTGWSDG-VCFILGLSTSCFMFIGL 256
Db 218 ---IYAIIFIDILLVFFHKGRSLKYALHFDNLLNSGYKSAPLSFIIGFQQSNFTLQGF 274
QY 257 DAAMHLAEECTDAARTVPKAVVSAIIIGFCTAFPTYIAVLYGITDLDLSILSSAGYIP--- 313
Db 275 SMLPALADEVKVPEKDIPRCMSNAVLISAFSGVIFLIPILMLILPDNDLLFTNHKVLPIVN 334

```
QY 314 -FETMTQSLRSLSFATVLSGCGIVMAFFALNAVQETASRLTWSFARDNGLVFSTHLRIH 372
Db 335 IFTKSTDSVLSFFLVLLILGNLL--FSGIGSI-TTSSRAVYSFSDQAIPIYDKWTYVE 391
QY 373 PRWQ--VPVMSLFATWILATCGCIFLGSSTAFAFNALVNSAVVLQQLSFLIPIALLLYOKR 430
Db 392 PDSQSKVPKNSVVLMSIISYFLGLLALISTAAFAFNAFAGAAVLCCLCSATFIPLVLFETR 451
QY 431 DPKFLPSTRAFPVLPGRIGFLVNVNAVVFSTVTVFFSFPLTVPTAASTNNY 481
Db 452 --RAIRAPVKIRYK-FGWFINIVSIWLLLSMVSVCCLPTQVPVTFKTMNY 499

RESULT 11
YCT3_SCHPO
ID YCT3_SCHPO STANDARD; PRT; 554 AA.
AC O59813; P78767;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative amino-acid permease C794.03.
GN SPCC794.03.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Murphy L., Harris D., Wood V., Rajandream M.A., Barrell B.G.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 198-554 FROM N.A.
RC STRAIN=PR745;
RX MEDLINE=98162722; PubMed=9501991;
RA Yoshioka S., Kato K., Nakai K., Okayama H., Nojima H.;
RT "Identification of open reading frames in Schizosaccharomyces pombe
cDNAs.";
RL DNA Res. 4:363-369(1997).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE AMINO ACID PERMEASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL023595; CAA19131.1; -
DR EMBL; D89115; BAA13777.1; -
DR InterPro; IPR002293; AA_rel_permease_1.
DR InterPro; IPR002027; Amino_acid_permease.
DR Pfam; PF00324; aa_permeases; 1.
DR PROSITE; PS00218; AMINO_ACID_PERMEASE_1; FALSE_NEG.
KW Hypothetical protein; Transport; Amino-acid transport; Transmembrane.
FT TRANSMEM 68 88
FT TRANSMEM 126 146
FT TRANSMEM 176 196
FT TRANSMEM 206 226
FT TRANSMEM 288 308
FT TRANSMEM 397 417
FT TRANSMEM 492 512
FT CONFLICT 293 293
FT CONFLICT 305 308
FT CONFLICT 313 313
FT CONFLICT 372 372
FT CONFLICT 412 412
FT CONFLICT 423 423
FT SEQUENCE 554 AA; 60857 MW; F88F245437F1265C CRC64;
SQ
```

```
Query Match 11.6%; Score 311.5; DB 1; Length 554;
Best Local Similarity 25.6%; Pred. No. 4.3e-13;
Matches 121; Conservative 88; Mismatches 221; Indels 43; Gaps 17;

QY 51 AFVICSWAGISGSLQALLAGGPVTLGLYILSTLVYICIAFSLAELTSVYPTAGGQYH 110
Db 54 SFAACDCMSNIRGSFYIGLLTGCPSAYWITYIIAIPQLISAATMAEVCALPTAGSLYF 113
QY 111 FASILAPKSNRSISYVCG--LVSLLSWIAIG-SSVTMIPAQQIPALIAAYSHYTSQDS 166
Db 114 WASAAGGKKYGRLLIGFIVAMVVVAVTWSFVAVNCQSTTKFIFGELPVNSGFSVSSDVK 173
QY 167 WHV--FLIYEGVALVLLNFALKNRPWVHVEIGFGLTIALFVIFISFIAL-LARSNPKAPN 223
Db 174 FRAVQWAVGEAILLVCLLNFIIPKWFYIFRSVAVILLDFVLNMWLPVAVSTKYGFR 233
QY 224 SQVWTAWSNY-----TGWSDGVCVFILGLSTSCFMFIFGLDAAMHLAEECTDAARTVPK-- 275
Db 234 DEAFMKSTNYDLGKVNNGWS----WCLTFFCTARILVGYDAAGHVAEETKNASKTASRGM 289
QY 276 ---AVVSAIL-IGFCTAPPYTIIVLYGIGITDLSILSSAGYIPFETMTQSLRSLSFATVLS 331
Db 290 FYSAFSNAILSTGIIVMFLYCLPPSNVMYELIKNSQQPFVSFAYALGKRAHVFMNVV- 348
QY 332 CGGIVMAFFALNAVQETASRLTWSFARDNGLVFSSTHLRIHPRWQVPVWLSLFATWGILAT 391
Db 349 --GILGMIFDTSLSIVASSRLVFAVARDGVLPFSGWLRKYVDSHGQ-PTNAVTFIFLISAA 405
QY 392 CGCIFLGSSTAFNALVSAVVLQQLSF-LIPIALLLYQKRDKPFLPS----TRAFVLP 445
Db 406 LLCNLPSSAVAFSTLSLAAAVPTIMAYAAVAFGRFLSLRNDPFPKSEWSLGLSKPFLQ-- 463
QY 446 GIGFLVNVNAVFTSVTVFFSFPLTVPTAASTNNYTSALIGVALGVLNVW 498
Db 464 -ITFLWN---LFTAV--ILFS-PKAYPVTGKNFNYAPVIFGAITIFGLISWL 508

RESULT 12
YFYA_SCHPO
ID YFYA_SCHPO STANDARD; PRT; 591 AA.
AC Q9UT18;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative amino-acid permease C9.10.
GN SPAC9.10.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Wedler H., Duesterhoeft A., Lyne M.H., Rajandream M.A., Barrell B.G.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE AMINO ACID PERMEASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL121764; CAB57428.1; -
DR InterPro; IPR002293; AA_rel_permease_1.
DR InterPro; IPR002027; Amino_acid_permease.
DR Pfam; PF00324; aa_permeases; 1.
```



```
DB 152 SVNLVANFTLFLVQLVISIMVVFIFLVVQGLHKGEGVGTWVWSLQPFISENAHLPIITGA 211
QY 247 STSCFMFIGLDAAMHLAEECTDAARTVPKAVYSAILIGFCTAFPTYTIAVLYGITDLDLSIL 306
DB 212 TIVCFSLGDFDAVTTLSEETPDAAARVIPKAI-----FLTAVYGGVIFIAASF 258
QY 307 SSAGYIP----FETMTQSLRSL-----PATVLSGGGIYMAFFALNAVQETASRLTWS 355
DB 259 FMQLFFPDISRFKDPDAALPEALYVGGKLFQSIFLCTTFVNTLASGLASHASVSRLLYV 318
QY 356 FARDNGLVFSTHLER----IHPRWQVPVWSLFWATWLGILATCGCIF-LGSSTA---FNLALV 407
DB 319 MGRDN--VFP---ERVFGYVHPKWRTPALNVIMV-GIVALSALFFDLVTATALINFGALV 372

RESULT 14
YEEF_ECOLI
ID YEEF_ECOLI STANDARD; PRT; 452 AA.
AC P33016;
DT 01-FEB-1994 (Rel. 28, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical transport protein yeeF.
GN YEEF OR B2014 OR Z3176 OR ECS2816.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN-K12 / BHB2600;
RA Richterich P., Lakey N., Gryan G., Jaehn L., Mintz L., Robison K.,
RA Church G.M.;
RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
RN [2]
SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [3]
SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=97251358; PubMed=9097040;
RA Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T.,
RA Isono K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T.,
RA Saito N., Sampei G., Seki Y., Sivasubraman S., Tagami H.,
RA Takeda J., Takemoto K., Wada C., Yamamoto Y., Horiuchi T.;
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 40.1-50.0 min region on the linkage map.";
RL DNA Res. 3:379-392(1996).
RN [4]
SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai J., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouisis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [5]
SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
```

```

MEDLINE=21156231; PubMed=11258796;
Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT *Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
CC -!- FUNCTION: PROBABLE AMINO-ACID OR METABOLITE TRANSPORT PROTEIN.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (Potential).
CC -!- SIMILARITY: BELONGS TO THE AMINO ACID PERMEASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U00009; AAA16420.1; ALT_INIT.
CC EMBL: AE000293; AAC75075.1; ALT_INIT.
CC EMBL: D90839; BAA15842.1; ALT_INIT.
CC EMBL: D90840; CAB21774.1; ALT_INIT.
CC EMBL: AE005427; AAG57073.1; ALT_INIT.
CC EMBL: AP002559; BAB36239.1; ALT_INIT.
CC EMBL: EGI1896; yeeF.
CC InterPro: IPR002293; AA_rel_permease_1.
CC InterPro: IPR002027; Amino_acid_permease.
CC Pfam: PF00324; aa_permeases; 1.
CC PROSITE: PS00218; AMINO_ACID_PERMEASE_1; FALSE_NEG.
KW Hypothetical protein; Transport; Amino-acid transport; Transmembrane;
KW Inner membrane; Complete proteome.
FT TRANSMEM 17 37 POTENTIAL.
FT TRANSMEM 49 69 POTENTIAL.
FT TRANSMEM 96 116 POTENTIAL.
FT TRANSMEM 124 144 POTENTIAL.
FT TRANSMEM 159 179 POTENTIAL.
FT TRANSMEM 200 220 POTENTIAL.
FT TRANSMEM 238 258 POTENTIAL.
FT TRANSMEM 284 304 POTENTIAL.
FT TRANSMEM 340 360 POTENTIAL.
FT TRANSMEM 361 381 POTENTIAL.
FT TRANSMEM 395 415 POTENTIAL.
FT TRANSMEM 418 438 POTENTIAL.
SQ SEQUENCE 452 AA; 49537 MW; F8B618C6BD3E762E CRC64;

Query Match 8.0%; Score 215.5; DB 1; Length 452;
Best Local Similarity 23.4%; Pred. No. 4.8e-07;
Matches 123; Conservative 74; Mismatches 182; Indels 147; Gaps 29;

QY 14 TROTTKNTEPAAAGGASESLNVPLEKKQFCFTIV-----SLAFV----ICNSWAGISGS 64
Db 2 SHNVTPNTS-----RVELRK---TLTLPVVMMLAYMQPMTLFDTFGIVSG- 45
QY 65 LQALLAGGPVTLTYGILISTLVYICIAFSLAELTSVYPTAGQYHFASILAPKSNRSI 124
Db 46 -----LTDGHVPTAYAFALIAILF--TALSCKLVRRYPSAGSAYTYAQ-----XSISPTV 94
QY 125 SYVCGVLSLSWIAIGSSVTMIPAAQIPALIAAAYSHYTSQDSWHVFLIYEGVALVLL-- 182
Db 95 GFMVGSLLDYLFAPMINILLAKIYFEALVP-----SIPSW-MFV-----VALVAFMTA 143
QY 183 FNLFAKRNPPWVHEIGFGLTIALFVISFTAI-----LARSNPKAPNSQ 225
Db 144 FNLRLSK-----SVANFNTVIVLVQVVLIAVLGMVVYGVFEGEGAGTLASTRP----- 192
QY 226 VWTAWSNYTGWSDGVCFILGLSTSCFMFIGLDAAMHLAEECTDAARTVPKAV-VSAIIIG 284
Db 193 ---FWS---GDAHVIPMITGATILCFSTFGDGISNLSSEETKDAERVIPRAIFLTALIGG 246
```

QY 285 PCTAPPYTIIVLYGTTDLSDLSILSSAG-----YIPFETMQSLRSLSFATVLSGCGIVM 337
Db 247 MIFIFATYFLQLY-FPDISRFKDPDASQPEIMLYVAGKAFQVGAALIFSTITVLSAG----- 301
QY 338 AFFALNAVQETASRLTWSFARDNGLVFSHTLEKIHPRQVVPVWSLFWATWGI----- 388
Db 302 -----MAAHAGVARLMVMYMGD-GVFPKSPFFGVHPKWRTPAMNIIYVGAIALAINFDL 355
QY 389 -LATCGCIFLGSSSTAFNALVNSAVVLQSLFSLPIALLLYQKRDPK-----FLPSTR 439
Db 356 VMAT-ALINFGALVAF-TFVNLSVISQ-----FWIREKRNKTLKDHFOYLFLEPCG 404
QY 440 AFVLPGRGIGEL-VN-----VLAVVFTSVTVFF-----SFPLTVP 473
Db 405 ALT-----VGALWVNLEESSMVLGLIWAAGIILYLACVTKSFRNPVP 446
RESULT 15
YJ99_MYCTU
ID YJ99_MYCTU STANDARD; PRT; 440 AA.
AC Q10858;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 45.3 kDa transport protein Rv1999c.
GN Rv1999C OR MT2055 OR MTCY39.19.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID-1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PROBABLE AMINO-ACID OR METABOLITE TRANSPORT PROTEIN.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE AMINO ACID PERMEASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; 274025; CAA98388.1; -.
DR EMBL; AE007057; AAK46332.1; -.
DR TIGR; MT2055; -.
DR TubercuList; Rv1999c; -.
DR InterPro; IPR002293; AA_rel_permease_1.

DR InterPro; IPR002422; AA_rel_permease_2.
DR InterPro; IPR002027; Amino_acid_permease.
DR Pfam; PF00324; aa_permeases; 1.
DR PROSITE; PS00218; AMINO_ACID_PERMEASE_1; FALSE_NEG.
KW Hypothetical protein; Transport; Transmembrane; Complete proteome.
FT TRANSMEM 24 44 POTENTIAL.
FT TRANSMEM 47 67 POTENTIAL.
FT TRANSMEM 93 113 POTENTIAL.
FT TRANSMEM 117 137 POTENTIAL.
FT TRANSMEM 155 175 POTENTIAL.
FT TRANSMEM 183 203 POTENTIAL.
FT TRANSMEM 229 249 POTENTIAL.
FT TRANSMEM 276 296 POTENTIAL.
FT TRANSMEM 323 343 POTENTIAL.
FT TRANSMEM 346 366 POTENTIAL.
FT TRANSMEM 379 399 POTENTIAL.
FT TRANSMEM 400 420 POTENTIAL.
SQ SEQUENCE 440 AA; 45267 MW; 9592EE0A842D18CD CRC64;
Query Match 7.4%; Score 199.5; DB 1; Length 440;
Best Local Similarity 23.9%; Pred. No. 5e-06;
Matches 117; Conservative 75; Mismatches 179; Indels 119; Gaps 25;
QY 34 NVPLE-KKQGTI--TIVSLAFVICSWAGISGLQALLAGQPVTLTYGILISTLVYIC 90
Db 9 DIPDELRRLGLLDVAVVIGLSMI--GAGIFAALAPAAVYAAGS-GLLLGLAFAAVVAYC 64
QY 91 IAFSLAELTSVYPTAGGQYHFASILAPKSIINRSISYVCGLVSLSW-IAICSSVTMIPAQ 149
Db 65 NAISSARLAARYPASGGTYVYGRM----RLGDFWGYLAG-----WGFVVGKT----- 107
QY 150 QIPALIAAYSHTYSQDSWHVFLIYEGVALVLL--FNLFALKRNPVW-HEICFGLTIALF 206
Db 108 ---ASCAAMALTGVGYVWPAQAHAHVAVAVVAVVAVVAVVAVVAVVAVVAVVAVVAVV 161
QY 207 VISFIALIARSNPKAP-----NSQVWTAWNSYTGWSGVCVFIGLSTSCFMFIGLDA 258
Db 162 VLTAVVVAAYGSGAADPARLDIGYDAHVW-----GMLQAAGL--LFFAFAGYAR 208
QY 259 AMHLAEECTDAARTVPKAVVSAIIIGFCTAFPTYIAVLYGITDLDLSILSSAGYIPFETMT 318
Db 209 IATLGEVRDPARTIPRAIPLALGI---TLAVYALVAVAVIAGVLPQRLARAAAP---LS 262
QY 319 QSLRSLS---FATVLSGCGIVMAFFALNAVQETASRLTWSFARDNGLVFSHTLER---I 371
Db 263 EAMRVACVNWLIIPVVOIGAAVAAALGSLALLILGVSRRTLAMARDR-----HLPRWLA 316
QY 372 HPRMQVP-----VWSLFATWGIATCGCIFLGSSSTAFNALVNSAVVLQSLFLIP 421
Db 317 HPRFKVPFRBELVVGAVVAALATADIRGAIG-----FSSFGVLVYVYAIANAS-----A 365
QY 422 IALLLYQKRDPKELPSTRAFLVLRGIGFLVNVAVVTSVTVFFSFPLTVPTAASTMNY 481
Db 366 LTLGLDEGRPRRLIP-----LVGLIGCV-----VLAFAFLPLSSVAAG----- 402
QY 482 TSAIIGVALA 491
Db 403 -AAVLGVGVA 411

Search completed: April 27, 2002, 07:59:41
Job time: 747 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2002, 07:48:21 ; Search time 68.15 Seconds
(without alignments)
740.233 Million cell updates/sec

Title: US-09-882-694A-8
Perfect score: 2692
Sequence: 1 MDSRPSGYGEKGGTRQTTKN.....QGPHELDGRVVGAFFQVGP 525

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	681.5	25.3	563	1 S11175	choline transport
2	540.5	20.1	561	2 S63388	probable membrane
3	454	16.9	567	2 T50051	amino-acid permeas
4	451	16.8	517	2 D84421	probable amino aci
5	445	16.5	542	2 T39474	amino acid permeas
6	429	15.9	557	2 T41588	probable amino-aci
7	420	15.6	544	2 S62527	probable amino-aci
8	383.5	14.2	571	1 S30253	GABA transport pro
9	351.5	13.1	550	2 T37519	probable amino aci
10	321	11.9	618	1 S38004	probable transport
11	311.5	11.6	554	2 T41612	amino acid permeas
12	274	10.2	591	2 T39195	probable amino aci
13	243.5	9.0	496	2 B83503	probable amino aci
14	227.5	8.5	465	2 E69825	amino acid transpo
15	226.5	8.4	479	2 C64878	probable amino aci
16	226.5	8.4	479	2 A90863	probable amino aci
17	225.5	8.4	461	2 D69814	metabolite transpo
18	225.5	8.4	461	2 H85755	probable amino aci
19	225	8.4	463	2 AH0189	probable amino aci
20	222.5	8.3	454	2 AG0763	probable amino aci
21	221	8.2	776	2 A84178	cationic amino aci
22	218	8.1	440	2 G83195	probable amino aci
23	216	8.0	482	2 C90067	hypothetical prote
24	215.5	8.0	454	2 E64966	probable amino aci
25	215.5	8.0	454	2 H90980	probable amino aci
26	215.5	8.0	454	2 E85826	probable amino aci
27	210.5	7.8	456	2 B83391	probable amino aci
28	202.5	7.5	494	2 B89827	hypothetical prote
29	201	7.5	463	2 AE1155	amino acid transpo

30	199.5	7.4	440	2 E70758	hypothetical prote
31	199	7.4	482	2 G83142	probable transport
32	199	7.4	556	2 G70503	probable cycA prot
33	197.5	7.3	596	2 S46001	probable amino aci
34	196	7.3	539	2 D69748	amino acid transpo
35	195	7.2	447	2 AI0701	probable amino aci
36	194	7.2	647	2 E90259	amino acid transpo
37	193	7.2	465	2 AG0415	aromatic amino aci
38	191.5	7.1	736	2 C69451	cationic amino aci
39	189.5	7.0	463	2 AH1758	amino acid transpo
40	189.5	7.0	496	2 G83136	probable amino aci
41	189.5	7.0	504	2 AG2839	hypothetical prote
42	189.5	7.0	518	2 A97617	probable transport
43	189	7.0	455	2 F83460	probable amino aci
44	189	7.0	465	2 A83046	probable amino aci
45	188.5	7.0	452	2 G95306	putrescine/ornithi

ALIGNMENTS

RESULT 1
S11175
choline transport protein - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein G3213; protein YGL077c
C;Species: Saccharomyces cerevisiae
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C;Accession: S11175; S64084
R;Nikawa, J.I.; Hosaka, K.; Tsukagoshi, Y.; Yamashita, S.
J. Biol. Chem. 265, 15996-16003, 1990
A;Title: Primary structure of the yeast choline transport gene and regulation of its
A;Reference number: S11175; MUID:90368823
A;Accession: S11175
A;Molecule type: DNA
A;Residues: 1-563 <NIK>
A;Cross-references: EMBL:J05603; NID:q171329; PIDN:AAA34537.1; PID:q171330
R;Rieger, M.; Mueller-Auer, S.; Brueckner, M.; Schaefer, M.
submitted to the Protein Sequence Database, May 1996
A;Reference number: S64071
A;Accession: S64084
A;Molecule type: DNA
A;Residues: 1-563 <RIE>
A;Cross-references: EMBL:Z72599; NID:q1322592; PIDN:CAA96782.1; PID:q1322593; GSPDB:Q
A;Experimental source: strain S288C
C;Genetics:
A;Gene: SGD:HNMI; CTRL: MIPS:YGL077c
A;Cross-references: MIPS:YGL077c; SGD:S0003045
A;Map position: 7L
C;Superfamily: choline transport protein
C;Keywords: transmembrane protein
F;91-107/Domain: transmembrane #status predicted <TM1>
F;185-201/Domain: transmembrane #status predicted <TM2>
F;214-230/Domain: transmembrane #status predicted <TM3>
F;257-273/Domain: transmembrane #status predicted <TM4>
F;298-314/Domain: transmembrane #status predicted <TM5>
F;345-361/Domain: transmembrane #status predicted <TM6>
F;401-417/Domain: transmembrane #status predicted <TM7>
F;428-444/Domain: transmembrane #status predicted <TM8>
F;466-482/Domain: transmembrane #status predicted <TM9>

Query Match 25.3%; Score 681.5; DB 1; Length 563;
Best Local Similarity 33.2%; Pred. No. 9.8e-42;
Matches 156; Conservative 95; Mismatches 210; Indels 9; Gaps 6;

Qy	39	KKQFGTITVSLAFVICNSWAGISGSLQLALLAGGPVTLTYGILISTLVYICIAFSLAEL	98
Db	55	RKSFSLWSILGVGGLTNSWFGISTSMVAGISSGGPMIVYGIITVALISICIGTSLGEL	114
Qy	99	TSVYPTAGQYHFASILAPKSIINRSISYVCGLVSLLSWIAIGSSVTMIPAOQIPALIAAY	158
Db	115	SSAYPHAGGQFWWSLKLAPPKYKRFAAYMCGSFAYAGSVFTSASTTSLSVATEVVGMYALT	174

C;Accession: T41588
R;Murphy, L.; Harris, D.; Lyne, M.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, September 1998
A;Reference number: Z21968
A;Accession: T41588
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-557 <MUR>
A;Cross-references: EMBL:AL031543; PIDN:CAA20834.1; GSPDB:GN00068; SPDB:SPCC74.04
A;Experimental source: strain 972h-; cosmid c74
C;Genetics:
A;Gene: SPDB:SPCC74.04
A;Map position: 3
C;Superfamily: choline transport protein

Query Match 15.9%; Score 429; DB 2; Length 557;
Best Local Similarity 24.1%; Pred. No. 1.7e-23;
Matches 130; Conservative 110; Mismatches 262; Indels 38; Gaps 11;

QY 3 SRPSGYGKGTQRTTKNTETAAGGASESLNVPLEKKQFGTITIVSLAFVICNSWAGIS 62
Db 27 SNPTNPEPNITSEADNAEDLAALGYKQEF-----QRGLSLFSVFSPLGLLPSVA 80

QY 63 GSLQLALLAGGPVTLVLYGILISTLVYICIAFSLAELTSVYPTAGQYHFASILAPKSINR 122
Db 81 TTLPSIGYGTGTPGLLWGLIAMVFIICIALSMAELCSAMPTSGGLYAAAVLAPEGWGP 140

QY 123 SISYVCGLSLLSWIAIGSSVTMIPAQQIPALIAAYSHYTSQDSWHVFLIYEGVALVLL 182
Db 141 LAAWFTGWSNYIAQLVGGPSINYSTAAMLLGAVNIGNPNYEVQNYQLFLVSIQFIHFI 200

QY 183 FNLFALKRNPWVHVEIGFGLTIALFVVISFIALIARSNPK--APNSQVWTAWSNYTGWSDG 239
Db 201 LASMPTKYIAKLNSVGTYLNTLFLFISMIVILAMSSKNHGFNETSKVWSHIENYTDWPDG 260

QY 240 VCFILGLSTSCFMFIGLDAAMHLAEECTDAARTVPKAVVSAIIIGFCTAFPPYTIADVLYGI 299
Db 261 FAILMSECGVIWMTSGYDAPFHMSEETANASVNAPRGIILTAAGINGMWVMQIVYATV 320

QY 300 TDLDSLSSAGYIPFETMTQSLRSLSPATVLSGCGIVMAFFALNAVQETASRLTWSFARD 359
Db 321 VDQTAVTGSDSMWATYLSQCLPKRAALGILSL-TIVSSFLMGQSNLIASRIAYSARD 379

QY 360 NGLVFSHLERIHPRWQVPVWSLFAWGI-----LATCGCIFLGSSTAFNALVNSAVVL 413
Db 380 GVLPSYSEWVATVNPITKTPIRAVFVNEVIGVLILFLAFAGAITIG-----AVFSVTAIA 433

QY 414 QQLSFLIPIALLLYQKRPKFLPSTRAFVL--PRGIGFLVNLAVVFTSVTVFFSFPL 470
Db 434 AFTAFVAPVAMRVFVKDADF--RTGPNLKGFSKPIGF---CSVSFVLMIPILCFPS 487

QY 471 TVPTAATMNYTSAIIGVALALGVLMW-VVHARKHYQGPLHL-----DGRVVGAEFQVG 524
Db 488 VKNPTPAEMNWTCLVFGAPM-LAVLIWYAISGRKWKFGKPRINLASSEGDNSTLEGVELYTG 546

RESULT 7
S62527
probable amino-acid permease - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 16-May-1996 #sequence_revision 13-Mar-1997 #text_change 10-Dec-1999
C;Accession: T41435; T39133; S62527
R;Seeger, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, October 1998
A;Reference number: Z21993
A;Accession: T41435
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-544 <SEE>
A;Cross-references: EMBL:AL032824; PIDN:CAB37426.1; GSPDB:GN00068; SPDB:SPCC584.13
A;Experimental source: strain 972h-; cosmid c584
R;Lye, G.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.

submitted to the EMBL Data Library, November 1995
A;Reference number: Z21830
A;Accession: T39133
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-544 <LYE>
A;Cross-references: EMBL:Z66569; NID:g1052533; PIDN:CAA91521.1; PID:g1052542; SPDB:S
C;Genetics: <SEE1>
A;Gene: SPCC584.13
A;Map position: 3
C;Genetics: <LYE2>
A;Gene: SPAC8A4.11
A;Map position: 1
C;Superfamily: choline transport protein

Query Match 15.6%; Score 420; DB 2; Length 544;
Best Local Similarity 24.1%; Pred. No. 7.4e-23;
Matches 120; Conservative 103; Mismatches 231; Indels 44; Gaps 11;

QY 39 KQFGTITIVSLAFVICNSWAGISGLALLAGGPVTLVLYGILISTLVYICIAFSLAEL 98
Db 38 KREFSAWTSFCVSFVLGLLPFSFASTMYTTGYAGTPAMVWGWLIAVFEVQCVANGMAEL 97

QY 99 TSVYPTAGQYHFASILAPKSINRSISYVCGLVSLLSWIAIGSSVTMIPAQQIPALIAAY 158
Db 98 CSSMPTSGGLYAAAVLAPKGGPFAAWLTGWSNYLVQVTGPPSVAYSPAGMILTIVQLH 157

QY 159 SHYTSQDSWHVFLIYEGVAL-----VLLFNLFALKRNPWVHVEIGFGLTIALFV 207
Db 158 NPFETQNYQIFLLAVAAMIAOQGFISSMPTKVLAVENTWGTVLNM-----LFLAIVM 209

QY 208 ISFIAILARSNPKAPNS--QVWTAWSNYTGWSDGVCFILGLSTSCFMFIGLDAAMHLAEE 265
Db 210 ITVLAVAGTKTPRGFNSNHKVNNEEDNOTDWSNGMAMLSFAGVIWMTSGYDSPFHLSEE 269

QY 266 CTDAARTVPKAVVSAIIIGFCTAFPPYTIADVLYGITDLSILSSAGYIPFETMTQSLRSL 325
Db 270 CSNASVAAPRAIVMTSAFGGIVGVLNLCIAYTIVDVNAAMNDLGGPPVYLRQVCNYK 329

QY 326 FATVLSGCGIVMAFFALNAVQETASRLTWSFARDNGLVFSHLERIHPRWQVP---VWSL 382
Db 330 TTVALTSLTVICSFMMGQGMVAASRVTSYVARDGVFPFSKYLAIVDKRTKTPNVCVW-M 388

QY 383 FATWGILATCGCIFLGSSTAFNALVNSAVVLQQLSFLIPIALLLYQKRPKFLPSTRAFV 442
Db 389 NVVVGILC-CLLIFAGEA-AINAIFSVGAIAAFVATTPIFLRVFFVKEDEFKRG----- 441

QY 443 LPRGIGFLVNV--LAVVFTSVTVFFSFPL-----TVPTAATMNYTSAIIGVALALGV 494
Db 442 -PWHLGKFSKINGYAAACAFVLLMVPILCFPPQFRGKDNTPDA---MNVTCVVFVGGPMLMVL 497

QY 495 LNVVHVHARKHYQGPLHEL 512
Db 498 IWMFVSARKWFKGPRLTI 515

RESULT 8
S30253
GABA transport protein - yeast (Saccharomyces cerevisiae)
N;Alternate names: GABA-specific permease; protein D1037; protein YDL210w
C;Species: Saccharomyces cerevisiae
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C;Accession: S30253; S67769; S25147
R;Andre, B.; Hein, C.; Grenson, M.; Jauniaux, J.C.
Mol. Gen. Genet. 237, 17-25, 1993
A;Title: Cloning and expression of the UGA4 gene coding for the inducible GABA-speci
A;Reference number: S30253; MUID:93204891
A;Accession: S30253
A;Molecule type: DNA
A;Residues: 1-571 <AND>
A;Cross-references: EMBL:X66472; NID:g4749; PIDN:CAA47101.1; PID:g4750
A;Note: the sequence from Fig. 5 is inconsistent with that from Fig. 3 in having 527

Db 109 GWDLLLEYMLAVSAVSGWGYFQSFSLGSLGIHLPV-ALTAAPGAVKGTFTLENLPAFVI 167
QY 173 YEGVALVLLFNLFAKRNPNVWHEIGFGLTIALFVIFIAILARSNPKNQVWTA--- 229
Db 168 V--MAITYLLY--LGIKESKRVNNIMVILKI-LVLLLFIAVAA-----VYVKPHN 212
QY 230 WSNYTGWSDGVCFILGLSTSCFMFIGLDAAMHLAEECTDAARTVPKAVVSAIIGECTAF 289
Db 213 WQPFMPMGFGGVF-SAAALVFFAFIGFDVAVSSAAEETKNPAKDLPGIIFSLV--CTIL 269
QY 290 PYTIAVLVYGITDLSILSSAGYIPF-----ETMTQSLRSLSFATVLSGCGIVMAF 339
Db 270 YVTVSAI-----MTGVIPFAQFAGVDHPVSLVLQSAQNWWAGIIDIGAVLGMT 318
QY 340 FALNAVQETASRLTWSFARDNGLVSTHLEIRHPRWQVPVWSLFSATWGIATCGCIFLGS 399
Db 319 TVMLVMLYGQTRVMFAMSRD-GLVPGS-LSKVHPKHKTP---YVATW-----FFGT 364
QY 400 STAFNALVNSAVVLOQLS-----FLIPIALLLYQKRDPKFLPSTRAFPVLPRGIG 448
Db 365 ---MSALLGSLVPLDELAKLVNIGTLSAFVLISAVIVLRKKQPD-LP--RAFKCP-GVP 417
QY 449 FLVNVLAIVFTSVTVTFSPFPLTVPTAASTMNYSATIGVALALGVLNVVHVHARKH 504
Db 418 -VIPGLAILE---CLFLILNLGWVTIVRFL-----VWLLIGLVIVFLYSRKH 460

RESULT 15

C64878
Probable amino acid permease ycjJ - Escherichia coli
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 17-Mar-2000
C:Accession: C64878
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: C64878
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-479 <BLAT>
A:Cross-references: GB:AE000227; GB:U00096; NID:g1787543; PIDN:AAC74378.1; PID:g1787553;
A:Experimental source: strain K-12, substrain MGL655
C:Genetics:
A:Gene: ymjJ
C:Superfamily: arginine permease
C:Keywords: amino acid transport; transmembrane protein
F:73-89/Domain: transmembrane #status predicted <TM1>
F:130-146/Domain: transmembrane #status predicted <TM2>
F:150-166/Domain: transmembrane #status predicted <TM3>
F:181-197/Domain: transmembrane #status predicted <TM4>
F:222-238/Domain: transmembrane #status predicted <TM5>
F:260-276/Domain: transmembrane #status predicted <TM6>
F:362-378/Domain: transmembrane #status predicted <TM7>
F:383-399/Domain: transmembrane #status predicted <TM8>
F:419-435/Domain: transmembrane #status predicted <TM9>

Query Match 8.4%; Score 226.5; DB 2; Length 479;
Best Local Similarity 22.1%; Pred. No. 6.3e-09;
Matches 100; Conservative 70; Mismatches 139; Indels 143; Gaps 18;

QY 25 AAGGASESLNVPL-----EKKQFGTITIVISLAFV----ICNSWAGISGSLQLA 68
Db 13 AGEAAIMAINSPNIAAQPGRKRLRKSCLKLWQVMMGLAYLTPMTVFDFPGIVSG----- 67
QY 69 LLAGGPVTLTYGILISTLVYICIAFSLAELTSVYPTAGQYHFASILAPKSINRSISYVC 128
Db 68 -ISDGHVPASVYLLALAGVLFYTAISY--GKLVQFPEAGSAYTYAQ----KSINPHVGFMV 120
QY 129 GLVSLLSWIAIGSSVTMIPAQQIPALIAAYSHYTSQDSWHVFLIIEGVALVLLFNLFAL 188

Db 121 GWSSLLDYL-----FLPMINVLLAKIYLSALPP 148
QY 189 KRNPNWVHEIGFGLTIALFVIFIAILARSNPKNAPN----- 223
Db 149 EVPPWV-----WVVTFAVAILTAANLKSUNLVANFNTLFLVLQVISIMVVFIFLVV 197
QY 224 -----SQVWTANSNYTGWSDG---VCFILGLSTSCFMFIGLDAAMHLAEECTDAARTVP 274
Db 198 QGLHKGEVGTVWSLQPPFISENAHLIP1ITGATIVCFSPFLGFDVAVTTLSEETPDAAARVIP 257
QY 275 KAVVSAIIIGFCTAFPYTIAVLVYGITDLSILSSAGYIP----FETMTQSLRSL- 325
Db 258 KAI-----FLTAVYGGVIFIAASFEMQLFFPDISRFKDPDAALPEIALYVGG 304
QY 326 --FATVLSGCGIVMAFFALNAVQETASRLTWSFARDNGLVSTHLE- ---IHPRWQVPV 379
Db 305 KLFQSIFLCTTTFVNTLASGLASHASVSRLLYVMGRDN--VFP---ERVFGYVHPKWRTPA 359
QY 380 WSLFATWGILATCGCIF-LGSSTA---FNALV 407
Db 360 LNVIMV-GIVALSALFFDLVTATALINFGALV 390

Search completed: April 27, 2002, 07:48:23
Job time: 6730 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2002, 07:58:37 ; Search time 24.91 Seconds
(without alignments)
147.422 Million cell updates/sec

Title: US-09-882-694A-8
Perfect score: 2692
Sequence: 1 MDSRPSGYGEKGCTROTCKN.....QGPLHLELDGRVVGAEFQVGP 525

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 40831 seqs, 6994805 residues

Total number of hits satisfying chosen parameters: 40831

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_AA_New:*
1: /cgn2_6/ptodata/2/paa/pct_NEW_COMB.pep:*
2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep:*
3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep:*
4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep:*
5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:*
6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:*
7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	171.5	6.4	714	1	PCT-US02-09944-473 Sequence 473, App
2	159	5.9	585	5	US-09-573-655B-1554 Sequence 1554, App
3	156	5.8	519	5	US-09-895-913A-118 Sequence 118, App
4	133.5	5.0	533	5	US-09-540-209B-6591 Sequence 6591, App
5	123	4.6	832	5	US-09-573-655B-1490 Sequence 1490, App
6	116.5	4.3	562	5	US-09-540-209B-5760 Sequence 5760, App
7	112	4.2	459	5	US-09-540-209B-6435 Sequence 6435, App
8	106.5	4.0	466	1	PCT-US02-09921-675 Sequence 675, App
9	106	3.9	483	5	US-09-573-655B-2098 Sequence 2098, App
10	105	3.9	456	1	PCT-US02-09962-17 Sequence 17, Appl
11	103.5	3.8	867	5	US-09-573-655B-364 Sequence 364, App
12	102	3.8	1089	6	US-10-006-063A-102 Sequence 102, App
13	102	3.8	1089	6	US-10-006-117A-102 Sequence 102, App
14	102	3.8	1089	6	US-10-006-130A-102 Sequence 102, App
15	102	3.8	1089	6	US-10-006-172A-102 Sequence 102, App
16	102	3.8	1089	6	US-10-006-768A-102 Sequence 102, App
17	102	3.8	1089	6	US-10-017-527A-102 Sequence 102, App
18	102	3.8	1089	6	US-10-017-610A-102 Sequence 102, App
19	100.5	3.7	429	5	US-09-540-209B-10419 Sequence 10419, A
20	100.5	3.7	450	5	US-09-540-209B-5857 Sequence 5857, App
21	98.5	3.7	488	6	US-10-109-310-15 Sequence 15, Appl
22	98.5	3.7	530	6	US-10-107-431-275 Sequence 275, App
23	97.5	3.6	466	5	US-09-540-209B-6619 Sequence 6619, App
24	97	3.6	515	5	US-09-540-209B-10338 Sequence 10338, A
25	96	3.6	340	6	US-10-108-605-291 Sequence 291, App
26	96	3.6	474	5	US-09-540-209B-6787 Sequence 6787, App

27	95.5	3.5	426	5	US-09-540-209B-5807	Sequence 5807, Ap
28	95.5	3.5	584	1	PCT-US02-09921-601	Sequence 601, App
29	95	3.5	445	6	US-10-107-431-91	Sequence 91, Appl
30	94	3.5	353	1	PCT-US02-05625-63	Sequence 63, Appl
31	93	3.5	498	5	US-09-540-209B-5414	Sequence 5414, Ap
32	93	3.5	836	5	US-09-540-209B-6015	Sequence 6015, Ap
33	92.5	3.4	442	1	PCT-US01-25881-26	Sequence 26, Appl
34	92.5	3.4	626	1	PCT-US02-07826-95	Sequence 95, Appl
35	92.5	3.4	626	6	US-10-097-340-95	Sequence 95, Appl
36	92	3.4	443	5	US-09-540-209B-9709	Sequence 9709, Ap
37	91	3.4	390	5	US-09-895-913A-286	Sequence 286, App
38	91	3.4	536	1	PCT-US02-09962-16	Sequence 16, Appl
39	90.5	3.4	430	5	US-09-540-209B-5690	Sequence 5690, Ap
40	90.5	3.4	462	6	US-10-108-605-345	Sequence 345, App
41	90.5	3.4	525	5	US-09-972-211-104	Sequence 104, App
42	89.5	3.3	521	5	US-09-972-211-106	Sequence 106, App
43	89.5	3.3	530	5	US-09-540-209B-7741	Sequence 7741, Ap
44	89.5	3.3	637	5	US-09-540-209B-5999	Sequence 5999, Ap
45	89	3.3	439	5	US-09-540-209B-5802	Sequence 5802, Ap

ALIGNMENTS

RESULT 1
PCT-US02-09944-473
; Sequence 473, Application PC/TUS0209944
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: DAFFO, Abel
; APPLICANT: JONES, Anissa L.
; APPLICANT: TRAN, Alanna-Phung B.
; APPLICANT: DAHL, Christopher R.
; APPLICANT: GIETZEN, Darryl
; APPLICANT: CHINN, Joyce
; APPLICANT: DUFOUR, Gerard E.
; APPLICANT: HILLMAN, Jennifer L.
; APPLICANT: YU, Jimmy Y.
; APPLICANT: TUASON, Olivia
; APPLICANT: YAP, Pierre E.
; APPLICANT: AMSHEY, Stefan R.
; APPLICANT: DAUGHERTY, Sean C.
; APPLICANT: DAM, Tam C.
; APPLICANT: LIU, Tommy F.
; APPLICANT: NGUYEN, Duy-Viet An
; APPLICANT: KLEEFELD, Yael
; APPLICANT: GERSTIN, JR., Edward H.
; APPLICANT: PERALTA, Careyna H.
; APPLICANT: DAVID, Marie H.
; APPLICANT: LEWIS, Samantha A.
; APPLICANT: CHEN, Alice J.
; APPLICANT: PANZER, Scott R.
; APPLICANT: HARRIS, Bernard
; APPLICANT: FLORES, Vincent
; APPLICANT: MARWAHA, Rakesh
; APPLICANT: LO, Audrey
; APPLICANT: LAN, Ruth Y.
; APPLICANT: URASHKA, Michael E.
; TITLE OF INVENTION: MOLECULES FOR DISEASE DETECTION AND TREATMENT
; FILE REFERENCE: PT-1231 PCT
; CURRENT APPLICATION NUMBER: PCT/US02/09944
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: 60/280,067; 60/279,619; 60/280,068; 60/291,280; 60/291,829; 60/299,428; 60/300,001; 60/299,776
; PRIOR FILING DATE: 2001-03-29; 2001-03-28; 2001-03-29; 2001-05-16; 2001-05-17;
; 2001-05-17; 2001-06-19; 2001-06-20; 2001-06-20
; NUMBER OF SEQ ID NOS: 792
; SOFTWARE: PERL Program
; SEQ ID NO 473
; LENGTH: 714
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:

```

; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: LG:199489.1.orf2:2001MAR30
PCT-US02-09944-473

```

Query Match 6.4%; Score 171.5; DB 1; Length 714;
Best Local Similarity 21.5%; Pred. No. 3e-06;
Matches 108; Conservative 88; Mismatches 199; Indels 107; Gaps 25;

Qy	66	QLALLAGPVTLLYGILISTLVYICIAFSLAELTSVYPTAGQYHFASTLAPKSINRSIS	125
Dd	114	EVAKADSGP-SIVVSFLIAALASVMAGLCYAEGFARVPKTGSAYLYTVV-----TVGELWA	168
Qy	126	YVCGLSVLLSWIAIGSSVTMIPAQQIPAL----IAAYSHTYSQDSWHVFLIYE--GVAL	178
Dd	169	FITGNWLILSYVIGTSSVARAWSGTFDLLSKQIGQFLRTYFRMNVTGLAEYPDFFAVCL	228
Qy	179	VWLLFNL--FALKRNPVHEI--GFGLTIALFV-----ISFIAILA	215
Dd	229	ILLLAGLLSFGVKESAWNKVFATAVNILVLFFVMVAGFVGKNVANWKISEEFLKNISASA	288
Qy	216	RSNPKAPNSQVWTAWS---NYTGWSGDGVCFILGLSTSCFMFIGLDAAAMHLAEECTDAAR	271
Dd	289	REPPSENGTSIYGAGFPMPYGFTG-----TLAGAATCFYAFVGDCIATTGEEVRNPQK	342
Qy	272	TVPKAVVSIIIGFCTAF---PYTIAVLYGITDLDSLSSA-----GYIPFETMTQSL	321
Dd	343	AIPIGIVTSLVFCFMAYEGVSAALTLMPPYYLLDEKSPLPVAFEVYCWGPAPKYVVAAAGSL	402
Qy	322	RSLSFATVLSGGIVMAFFALNAVQETASRLTWSFARDNGLVFSTHLERIHPRWQVPVWS	381
Dd	403	CALSTS-----LLGSIFPM-----PRVIYAMAED-GLLFKC-LAQINSKTKTPIIA	446
Qy	382	LFATWGILATCGCIFLGSSTA FNALVN--SAVVLOQLSFLIPTALLL-YQK----RDPKF	434
Dd	447	TLSSGAVAALMAFLF-----DLKALVDMMSIGTLMAYSLVAACVLIRYQPGLSYDQPKC	501
Qy	435	LPSTRAFVLPRGIGFLNVNLAVVFTSVT-----TVFFSPFLTVPVTAASTMNYTSA	484
Dd	502	SPE-----KDGLGSSPRVTSKSESQVMTLOROGFSMRTLF--CPSSLPTQOSA-SLVSF	552
Qy	485	IIG-----VALALGVL-NWVVHA	501
Dd	553	LVGFLAFLVLGLSVLTITYGVHA	574

RESULT 2
 US-09-573-655B-1554
 ; Sequence 1554, Application US/09573655B
 ; GENERAL INFORMATION:
 ; APPLICANT: SOLOVYEV, Victor and TROUKHAN, Maxim
 ; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
 ; TITLE OF INVENTION: Thereby
 ; FILE REFERENCE: 2750-0876P
 ; CURRENT APPLICATION NUMBER: US/09/573,655B
 ; CURRENT FILING DATE: 2000-05-18
 ; NUMBER OF SEQ ID NOS: 3281
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 1554
 ; LENGTH: 585
 ; TYPE: PRT
 ; ORGANISM: Arabidopsis thaliana
 US-09-573-655B-1554

```

Query Match          5.9%; Score 159; DB 5; Length 585;
Best Local Similarity 21.4%; Pred. NO. 2.2e-05;
Matches 114; Conservative 83; Mismatches 170; Indels 166; Gaps 28;

```

Qy	42	FGTITVSLAFVICNSWAGISGSLQALLAGGPVFLYGI-----LISTLVYICIAFSLA	96
	:: : :		
Db	90	FGSV-VGSGVFVITGQEARVG-----AGPAVVLSYAISGVSAALLSVLCY-----A	133

QY	97	ELTSVYPTAGQYHFASILAPKSI	NRISYVCGLVSLLSWIAIGSSV--	TMIPAAQIPAL	154
Db	134	EFGEIPVAGGSFYL	RV-----ELGDFIAFIAAGNILL	EA	179
QY	155	IAAY-SHTYSQDSWHV-----	FLIYEGVALVLL-----	FNLFALKRNPWVHEIGF	199
Db	180	WSSYLASLVKNDSDYFRIKVD	SPAKGFDLLDPVAVAVLLVANGI	AMTGTGRTSWNLITS	239
QY	200	GLTIALFVISFIAILARS	NPKNPNSQVWTAWSNYTGWSDG	VCFIILGLSTSPFMEIGLDA	259
Db	240	MVTVCIIV--FIVVGFTH	SKTSNLPVFFPYG-----	AKGV--VQSAVVVWSYTG	290
QY	260	MHLAEECTDAARTVPKAVV---	SALIIIGFC-TAFPYTIAVLYG	ITDLDLSILSSAGYIPFE	315
Db	291	ANMAEETEKPSRDIPIGLV	GSMTVVYCLMALALTMVKY--	TEID--ANAAY----	341
QY	316	TMTQSLRSLSFATV-LSCGG	IVMAFFALNAVQETASRLTWS	FARDNGLVFSTHLER----	370
Db	342	-----SVAFAQIGMKWAK	YLVGICALKGM--TTSLLVGS	LQAR---YTQIARSHMI	389
QY	371	-----IHPRQVPVWSL-----	-----FATW-----	-----	386
Db	390	PPWFALVHPKGTPIYATLL	VTLSSIIISFPTSLEVLSSV	FSFATLFIEMLVAVALLVRR	449
QY	387	-----GILATCGCIFL--	GSSTAFNALVNSAV-----	VLQQLSFLIPIALLLY	427
Db	450	YVVKDVTPEAGLLKFLFL	FLIIASSIGVSALWNSGVK	GWIAYTVTGVIWFIGTLGL	509
QY	428	QK-RDPK-----FLPS---	TRAFVLPRGIGFLVNLAV	VFTSVTTVFFSF	468
Db	510	PKYRVPKVGVPVLPWLP	SFSSTAMNLFILGSLGYA	FLRFIICTMVMLLYYLF	562
RESULT 3					
US-09-895-913A-118					
; Sequence 118, Application US/09895913A					
; GENERAL INFORMATION:					
; APPLICANT: Kleanthous, Harold					
; APPLICANT: Al-Garawi, Amal					
; APPLICANT: Miller, Charles					
; APPLICANT: Tomb, Jean Francois					
; APPLICANT: Oomen, Raymond P.					
; TITLE OF INVENTION: Identification of Polynucleotides					
; TITLE OF INVENTION: Encoding Novel Helicobacter Polypeptides in the He					
; FILE OF INVENTION: Genome					
; FILE REFERENCE: 06132/043002					
; CURRENT APPLICATION NUMBER: US/09/895,913A					
; CURRENT FILING DATE: 2001-06-29					
; PRIOR APPLICATION NUMBER: US 08/881,227					
; PRIOR FILING DATE: 1997-06-24					
; NUMBER OF SEQ ID NOS: 368					
; SOFTWARE: FastSeq for Windows Version 4.0					
; SEQ ID NO 118					
; LENGTH: 519					
; TYPE: PRT					
; ORGANISM: Helicobacter pylori					
US-09-895-913A-118					

Query Match 5.8%; Score 156; DB 5; Length 519;
Best Local Similarity 21.6%; Pred. NO. 3.2e-05;
Matches 108; Conservative 79; Mismatches 181; Indels 132; Gaps 23;

Qy	58	WAGISGSLQLALLAGGPVTLGYILISTLVYICIAFSLAELTSVYP	TAG--GOYHFASIL	115
Dd	80	FVGTCGNIAASA---GPLGLTIAYCFGLVWVCIMLSL	GELASVYPTTGSFGDY-----	129
Qy	116	APKSINRSISYVCGVLSSLWSIAIGSSVTMTIPAQQIPALIAAYSHTYSQDSWHVF	LIVYG	175
Dd	130	AAKFIGPGTGYMFWMYLGVW----ITV--ALEYIAIGMLMQRFADIPHYWVIL-C	181	
Qy	176	VALVVLLFNLFALKRNPWVHEICFGLTIALFVISFIAILARSNPKAPNSQVWTAWSNYTG	235	

Db 182 IALVELL-NFFSVKIFA-EGEFFSLIKVLAVIAFIGIGA-----IGIIYQIYSHGFG 232
QY 236 -----WSDGVCFIL-GLSTSCFMFIGLDAAMHLAECTDAARTVPKAVUSAI 281
Db 233 SIFDNFHDGKGFPPNGSAAVFSAMLAVIFAFTGTEVICGAVGETKNASEVMPKAIKATL 292
QY 282 --IIGFCTAFPTYTIAVLYGINDLSILSSAGYIPFETMTQSLRSLF-----ATVLS 331
Db 293 WRIVEFFLGSVFISVFLPMND-----SSITQSPFVSFLERI-NLPFIGMGIYPVADIMN 346
QY 332 CGGIVMAFFALNAVQETASRLTWSFARDNGLVFSTHLERIHPRWQVPVWSL FATWGILAT 391
Db 347 AVIITAMFSTANSGLYGASRMIIYGLSKQK--MEFKVFSQLN-RQGTPTYAMFFS----- 397
QY 392 CGCIFLGSSTAFNALVNSAVVLOQLSF-LIPIALLLYQKRDPKFLPSTRAFVLP RGIGFL 450
Db 398 -----LSFSLIGLLVQIYAKENV-----VEAL 419
QY 451 VNULA---VVFTSVTTVFFSFPL-----TVPTAASTMNYTSAIIGVALALGVLN 496
Db 420 INVISFTVILVWVSVSQYSEFRKQYLKAGHSLEDLPYKAPFLPQLIGITGCAIGVI- 478
QY 497 WVHARKHYQGPHELDGRV 516
Db 479 -----GSAMDKDQRI 488

RESULT 4

US-09-540-209B-6591
; Sequence 6591, Application US/09540209B
; GENERAL INFORMATION:

; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES FRA
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1001-001
; CURRENT APPLICATION NUMBER: US/09/540,209B
; NUMBER OF SEQ ID NOS: 2000-04-04
; SEQ ID NO 6591

; LENGTH: 533
; TYPE: PRT
; ORGANISM: B.fragilis
US-09-540-209B-6591

Query Match 5.0%; Score 133.5; DB 5; Length 533;
Best Local Similarity 20.7%; Pred. No. 0.0018;
Matches 106; Conservative 105; Mismatches 201; Indels 101; Gaps 28;

QY 40 KQFGTTIVSLAFVICNSWAGISGSLQALLAGGPVTLTYGI-----LISTLVYICIAF 93
Db 6 KQAVKLGVFTLA--IMNVTAVVS-----LRGLPAEAVYGMSSAFYFLFAAIVFL-IPT 55
QY 94 SL--AELTSVY-PTAGGQYHFASILAPKSNRSISYVCGLVSLLSW----IAIGS-SVTM 145
Db 56 SLVAAELAAAFQDKQGGVFRWVGEAYGKKLGFLAIWVQWIESTI-WYPTVLTFFGAVSIAF 114
QY 146 IPAQOIPALIAAYSHTYSDSWHVFLIYEGVALVLLFNLFALKRNPWVHE-----IG 198
Db 115 I-----GMNDTHDMTLASNKYTYTLAV---VLIYYWLATFISLKGMGWGVKAKIGGMVG 165
QY 199 FGLTIALVFISFIAILA---RSNPKAPNSQVWTAWSNYTGSDGVCFCILGISTCFMF-I 254
Db 166 TIIPAALLIILGIVYLASGGHSNLDH-HSFFFPDLTNF-----DNVV-----LAASIFLFYA 216
QY 255 GLDAAMHLAECTDAARTVPKAV-----VSAIIGFCTAFPTYTIAVLYGITDLSILSSA 309
Db 217 GMEGGIHKVDMQNPSKNYPKAVFIGALITVILFVLGT---FSLGIITIPAKDIS--LTQS 271
QY 310 GIYPFETMTQSLRSLSFATVLSGCGIVMAFFALNAVQETASRLTWSFARDNGLVFSTHLE 369
Db 272 LLVGFDNFRYIHASWLSPIIA---IALAFGVLAGV-----LTWVAGPSKGIFAVGKAG 322

QY 370 RIHPRWQ-----VPVWSLFATWIGILATCGCIFL---GSSTAFNALVNSAVVLOQLSFLI 420
Db 323 YMPFFFOKTNKLGQVKNILFVQCGAVTVLSLLFVVMPSVQSFYQILSOLTIVILVVMYLL 382
QY 421 PIALLLYQKRDPKFLPSTRAFVLP-----GIGFLVNLVAVVFTSVTTVFFSFPL 470
Db 383 MFSGAIYLRNKM--KANRPRFRIGKKGNGLMWIVGGGLGSLLAFLS-----FIPPS 434
QY 471 TVPTAASTMNYTSAIIGVALALGVLNWVHVHARK 503
Db 435 QISTGSNTVWFSVLIIG-ALVVVIAPFIIYAAK 466

RESULT 5

US-09-573-655B-1490
; Sequence 1490, Application US/09573655B
; GENERAL INFORMATION:

; APPLICANT: SOLOVYEV, Victor and TROUKHAN, Maxim
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptid
; TITLE OF INVENTION: Thereby
; FILE REFERENCE: 2750-0876P
; CURRENT APPLICATION NUMBER: US/09/573,655B
; NUMBER OF SEQ ID NOS: 3281
; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 1490
; LENGTH: 832
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-573-655B-1490

Query Match 4.6%; Score 123; DB 5; Length 832;
Best Local Similarity 21.0%; Pred. No. 0.019;
Matches 106; Conservative 79; Mismatches 178; Indels 142; Gaps 29;

QY 26 AGGASESLNVPLEKKQFGTITIVSLAFVICNSWAGISGSLQALLAGGP-----VTL 77
Db 2 ADPAAESIDA--SSSRFGR-----VVCYNQTAMHGSNTIS--AAAPFFMTQLSVANL 49
QY 78 LYGILISTLVYICIAFSLAELTSVYPTAGGQYHFASILAPKSNRSISYVCGLVSLLSWI 137
Db 50 TYRILYYFLKPLCLPPFVAQI-----LCGL--LFSPT 79
QY 138 AIGSSVTMIPAQQIIPALIAAYSHTYSDSWHVFLIYEGVALVLLFNLFALKRNPWVHEI 197
Db 80 VLGNNNEWVL-----KLIFPYKYT-----MLLETFANLALVYNVFL----- 115
QY 198 GFGLTIALFVIS-----FIAILARSNPKAPNSQVWTAWSNYTGSDGV---CFILGLSTS 249
Db 116 GLGLDLRMKIKDKIPVIAIIVGLLAALLAGAGLYYLPNS--GEADKILAGCMYWSIAFG 173
QY 250 CFMFIGLDAAMHLAE--ECTDAARTVPKAVVSAIIIGFCTAFPTYTIAVLYGITDLSIL 306
Db 174 CTNF--POLARILADLKLRTDMGHT--AMCAAVVTDLCT---WILFIFGM---AIF 220
QY 307 SSAGYIPFETMTQSLRSLSFATVLSGCGIVMAFFALNAVQETASRLTWSFAR--DNGLVF 364
Db 221 SKSG-VRNEMLPYSLAS-TIAFVLLC-----YFVIQ-----PCGVAFIENNTVEGGQVG 266
QY 365 STHLERIHPRWQ-----VPVWSLFA-TWIGILATCGCIFLGSSTAFNALVNSAVVLOQLSFL 419
Db 267 DTHV-----WYTIAGVVICSLITEVCGVHSITGAFLEGLSIPHDIIRKMIIEKHLHDFL 320
QY 420 IPIALLLYQ-----KRDPKFLPSTRA---FVLPRGIGFLVNLVAVVFTSVTTVFFSFPLT 471
Db 321 SGMLMPLFYIICGLRADIGYMNRTVSGMMAVVTASVMVKILSTMFCS---IFLRIPLR 377
QY 472 VPTAASTMNYTSAIIGVALALGVLN 496
Db 378 DGLAIGALMNTKG---TMALVILN 398

QY	145	MIPAAQIPALIAA	YSH	TSQDS	WHVFLI	YEGVAL	VVLLFN	LNFALK	RNP-W	-----	VHEIG	199																																											
Db	197	TIVYVWMFALL	AYVM	TSR	-----	LKSIN	VFG	YHKPQW	KAQVRL	FKLG	241																																												
QY	199	FGLTIALF	--VIS	FIALARS	NP	KAP	-----	NSQV	TAW	SNYT	GSDG	VC	FILGL	STSC	FM	252																																							
Db	242	FPVAALFF	EV	TLFA	VVALV	SP	LP	II	IAA	HQ	VA	INF	SS	-----	LV	FMLP	MS	VGA	--	292																																			
QY	253	FIGLDAAM	H	LAEE	CTDA	ARTV	PKAV	VS	AI	IIG	FCTA	FP	Y	TI	AV	LY	G	ITD	LS	S	AGY	I	312																																
Db	293	AVSIRV	GH	R	LG	EE	N	V	D	G	AR	V	AS	R	V	-----	G	IM	V	G	L	A	---	T	IT	IT	V	LS	R	E	L	I	A	E	L	Y	T	N	346																
QY	313	PFETMTQ	S	RLS	F	AT	V	L	S	C	G	I	V	M	A	F	F	A	L	N	A	V	Q	E	T	A	S	R	L	T	W	S	F	A	R	D	N	G	L	V	F	S	T	H	L	E	R	I	H	372					
Db	347	P-EVIT	L	A	M	Q	L	L	F	A	A	V	Q	C	-----	T	D	A	V	Q	V	I	A	A	G	A	L	R	G	Y	-	K	D	M	R	A	I	F	N	R	-----	389													
QY	373	PRQVVP	V	W	S	L	F	A	T	W	G	I	L	A	T	C	G	I	F	L	S	S	T	A	F	N	A	L	V	N	S	A	V	V	L	Q	L	S	P	L	I	P	I	A	L	L	Y	Q	K	R	D	392			
Db	390	-----	T	F	I	A	Y	W	-----	I	L	G	L	P	T	G	Y	-----	403																																				
QY	433	KELP	S	T	R	A	F	V	L	P	R	G	-----	I	G	F	L	V	N	V	L	A	V	V	E	T	S	V	T	T	V	F	F	S	E	P	L	T	V	P	T	A	A	S	T	M	N	Y	T	S	A	I	I	G	487
Db	404	-ILGR	T	D	W	I	V	E	P	M	G	A	Q	G	F	W	L	G	F	I	I	G	L	T	A	-----	430																												
QY	488	VALALGV	-L	N	W	V	H	A	R	K	H	Y	Q	G	P	H	L	E	D	513																																			
Db	431	AALMLGV	R	L	R	W	M	-----	H	R	Q	E	P	D	V	Q	L	N	452																																				

RESULT 11

US-09-573-655B-364

; Sequence 364, Application US/09573655B

; GENERAL INFORMATION:

; APPLICANT: SOLOVYEV, Victor and TROUKHAN, Maxim

; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding

; TITLE OF INVENTION: Thereby

; FILE REFERENCE: 2750-0876P

; CURRENT APPLICATION NUMBER: US/09/573,655B

; CURRENT FILING DATE: 2000-05-18

; NUMBER OF SEQ ID NOS: 3281

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 364

; LENGTH: 867

; TYPE: PRT

; ORGANISM: Arabidopsis thaliana

US-09-573-655B-364

```
Query Match          3.8%; Score 103.5; DB 5; Length 867;  
Best Local Similarity 21.1%; Pred. NO. 0.61;  
Matches 108; Conservative 67; Mismatches 177; Indels 159; Gaps 27;
```

QY	34	NVPLEKKQFGTITIV--SLAFVICNSWAGISGSQLALLAGGPVTLLYGILISTLVYICI	91
	:	: : : :	:
Dd	13	NYDIEESRFGKIVCYDQSLLFEKREKGWEGS-----TLASSLPFFFIT	56
	:	: : : :	:
QY	92	AFLAEALT--SVYPTAGGYHFASILAPKSINRSISVVCGLVSLLSWAIGSSVTMIPAQ	149
	:	: : : :	:
Dd	57	QLFVANLSTRVLV----YLTRPLYLPPFY---AQILCGL--LFSPSVLGNTRFIIA--	103
	:	: : : :	:
QY	150	QIPALIAAYSHTYSQDSWHVF-----LIYEGVALVVLLFNLFAL-----KRNP	192
	:	: : : :	:
Dd	104	-----HVFPYRTMVLETFANALAVYNIFLLGLCMDLRVMVRITELKP	145
	:	: : : :	:
QY	193	VWHEIGF-GLTTIALFVISFIAILARSNPKAPNSQVWTAWSNYTGWSDGV---CFILGLST	248
	:	: : : :	:
Dd	146	VI--I AFTGLLVALPVGAFLYYL-----PGN-----GHDPDKIISGCVFWSVAL	186
	:	: : : :	:
QY	249	SCFMFIGLDAAMHLAE---ECTDAARTVPKAVVSATIIIGFCTAFPYTIAVLVGITDLDSI	305
	:	: : : :	:
Dd	187	ACTNF--PDLARILADLKLLSRDMGMT---AMCAALVTDLT---WVLLVFEGFASEFSKS	237
	:	: : : :	:

```

QY 306 LSSAGYIPFETMTQSLRSLSPATVLSLGGGIVMAFFALNAVQETASRLTWSFAR--DNGLV 363
Db 238 GTWNKMPFVIIT-----TAIFVLLCIFVIRPGIAWIFAKTVKAGHV 279
QY 364 FSTHLERIHPRWQVPVWSLFA-----TWGILATCGCIFLGSSTAFNALVNSAVVL 413
Db 280 GDTH-----VWFILGGVVLCGLITDACGVHSITGAFLFGLSIPDHIIIRNMIEE 328
QY 414 QOLSFLIPIALLLYQ-----KRDPKF-LPSTRAFLVLRGI--GFLVNVLAUVFTSVTTVF 465
Db 329 KLHDFLSGILMPLFYIICGLRADIGFMLQFTDKFMNVVVICSSFLVK---IVTTVITSLF 385
QY 466 FSFPLTVPTAASTMNNTSIIIGVALALGVLN 496
Db 386 MHIPMRDAFAIGALMNTKG----TSLSLVVLN 412

RESULT 12
US-10-006-063A-102
; Sequence 102, Application US/10006063A
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan I.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: P2830P1C3
; CURRENT APPLICATION NUMBER: US/10/006,063A
; CURRENT FILING DATE: 2002-03-15
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 477
; SEQ ID NO 102
; LENGTH: 1089
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-006-063A-102

```

```

Query Match      3.8%; Score 102; DB 6; Length 1089;
Best Local Similarity 21.0%; Pred. NO. 1;
Matches 99; Conservative 54; Mismatches 177; Indels 142; Gaps 26;

QY 25 AAGGASESNVPLEKKQFGTITIVS--LAFVICNSWAGTSGSLQLALL--AGGPVTLTYG 80
    | | : : : | : | | : : | | : | | | |
Db 496 AYAGLLGTIELKLDLVLLGAVAAVSSELPF-LWKAWAGWGSKRPLATLEPIGPVLL-- 552
    | | : : : | : | | : : | | : | | | |

QY 81 ILISTLVYICIAFSLAELTSVYPTAGG-----QYHF-ASILAPKSIINRSISYVCGLV 132
    | | : : | | : | | : : | | : | | | |
Db 553 LLFRLAVFFSDSFVVAEARATPFLGSGFILLVLVQLHWEGQLLPK----- 598
    | | : : | | : | | : : | | : | | | |

QY 133 LLSWIAIGSSVTMIPAQQIPALIAAYSHTYSQDSWHVFLIYEGVALVLLFNLF----- 186
    | | : | | | | | | | : : : | | : | | | |
Db 599 LLTMPRLGTSATNP-----PRHNGAYA-----LRLGIGLLLLCTRLAGLFHRCPEE 644
    | | : | | | | | | | : : : | | : | | | |

QY 187 --ALKRNPVWHEIG-----FGLTIALFVISFTAI---LAR-SNPKAPN-SQVWT 228
    | | : | | : | | : | | : | | : | | : | | : | |
Db 645 TPVCHSSPWLSPLASNMVGGRAKNLWYGACVAALVALLAAVRLWLRRYGNLKSPEPPMLFV 704
    | | : | | : | | : | | : | | : | | : | | : | |

QY 229 AWSNYTGWSDGVCFFILGLSTSCFMFI--GLD-AAMHIAEECTDAARTVPKAVVSAIIIGF 285
    | | : : | | : | | : | | : | | : | | : | | : | |
Db 705 RWG-----LPLMALGTAAAYWALASGADEAPPRRLVLYVSGASMWLPRVAVGLAASGL 755
    | | : : | | : | | : | | : | | : | | : | | : | |

```

```
QY 286 CTAF--PYTIAVLYGI-----TDLDSILSSAGYIPFETMTQSLRSL 324
Db 756 ALLLWKPVTLVKAGAGAPRTRTTLTPFSGPPTSQADLDYVVPQI-----YRHMQEERFGR- 810
QY 325 SFATVLSGGIVMAFFALNAVQETA--SRLTWSFARDNGLVFS---THLERIHPRW----- 375
Db 811 RLERTKSGPLTVAAYQLGVSYSAMVTALT-----LAFPLLLLHAERISLVFLLLF 863
QY 376 -----QVPV-----WSLFATWGIATCGCIFLGSSSTAFNAL 406
Db 864 LQSFLLHLLAAGIPVTTTPGPFTVPWQAVSAWALMATQFFYSTGHQPVFPPI 915

RESULT 13
US-10-006-117A-102
; Sequence 102, Application US/10006117A
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan I.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: P2830P1C13
; CURRENT APPLICATION NUMBER: US/10/006,117A
; CURRENT FILING DATE: 2002-03-19
; Prior Application removed - See File Wrapper or Palm
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 477
; SEQ ID NO 102
; LENGTH: 1089
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-006-117A-102
```

```
Query Match 3.8%; Score 102; DB 6; Length 1089;
Best Local Similarity 21.0%; Pred. No. 1;
Matches 99; Conservative 54; Mismatches 177; Indels 142; Gaps 26;

QY 25 AAGGASESLNVPLEKKQFGTITIVS--LAFVICNSWAGISGSLQLALL--AGGPVTLLYG 80
Db 496 AYAGLLGTIELKLDLVLLGAAVSSFLPF-LWKAWAGWGSKRPLATLFPPIGPVLL-- 552
QY 81 ILISTLVYICIAFSLAELTSVYPTAGG-----QYHF-ASILAPKSINRSISYVCGLVS 132
Db 553 LLFRLAVFFSDSFVVAERATPFLGSGFILLVVQLHWEQQLPPK----- 598
QY 133 LLSWIAIGSSVTMIPAAQIPALIAAYSHYTSQDSWHVFLIYEGVALVLLFNLF----- 186
Db 599 LLTMPRLGTSATNP-----PRHNGAYA-----LRLGIGLLCTRLAGLFHRCPEE 644
QY 187 --ALKRNPWVHEIG-----FGLTIALFVISFIAI---LAR-SNPKAPN-SQVMT 228
Db 645 TPVCHSSPWLSPASMVGGRAKNLWYGACVAALVALLAAVRLWRLRRYGNLKSPEPPMLFV 704
QY 229 AWSNYTGWSDGVCFILGLSTSCFMFI--GLD-AAMHLAECTDAARTVPKAVVSAIILGF 285
Db 705 RWG-----LPLMALGTAAYWALASGADEAPPRLRLVYSGASMLPRAVAGLAASGL 755
QY 286 CTAF--PYTIAVLYGI-----TDLDSILSSAGYIPFETMTQSLRSL 324
Db 756 ALLLWKPVTLVKAGAGAPRTRTTLTPFSGPPTSQADLDYVVPQI-----YRHMQEERFGR- 810
```

```
QY 325 SFATVLSGGIVMAFFALNAVQETA--SRLTWSFARDNGLVFS---THLERIHPRW----- 375
Db 811 RLERTKSGPLTVAAYQLGVSYSAMVTALT-----LAFPLLLLHAERISLVFLLLF 863
QY 376 -----QVPV-----WSLFATWGIATCGCIFLGSSSTAFNAL 406
Db 864 LOSFLLHLLAAGIPVTTTPGPFTVPWQAVSAWALMATQFFYSTGHQPVFPPI 915

RESULT 14
US-10-006-130A-102
; Sequence 102, Application US/10006130A
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan I.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: P2830P1C7
; CURRENT APPLICATION NUMBER: US/10/006,130A
; CURRENT FILING DATE: 2002-03-19
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 477
; SEQ ID NO 102
; LENGTH: 1089
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-006-130A-102
```

```
Query Match 3.8%; Score 102; DB 6; Length 1089;
Best Local Similarity 21.0%; Pred. No. 1;
Matches 99; Conservative 54; Mismatches 177; Indels 142; Gaps 26;

QY 25 AAGGASESLNVPLEKKQFGTITIVS--LAFVICNSWAGISGSLQLALL--AGGPVTLLYG 80
Db 496 AYAGLLGTIELKLDLVLLGAAVSSFLPF-LWKAWAGWGSKRPLATLFPPIGPVLL-- 552
QY 81 ILISTLVYICIAFSLAELTSVYPTAGG-----QYHF-ASILAPKSINRSISYVCGLVS 132
Db 553 LLFRLAVFFSDSFVVAERATPFLGSGFILLVVQLHWEQQLPPK----- 598
QY 133 LLSWIAIGSSVTMIPAAQIPALIAAYSHYTSQDSWHVFLIYEGVALVLLFNLF----- 186
Db 599 LLTMPRLGTSATNP-----PRHNGAYA-----LRLGIGLLCTRLAGLFHRCPEE 644
QY 187 --ALKRNPWVHEIG-----FGLTIALFVISFIAI---LAR-SNPKAPN-SQVMT 228
Db 645 TPVCHSSPWLSPASMVGGRAKNLWYGACVAALVALLAAVRLWRLRRYGNLKSPEPPMLFV 704
QY 229 AWSNYTGWSDGVCFILGLSTSCFMFI--GLD-AAMHLAECTDAARTVPKAVVSAIILGF 285
Db 705 RWG-----LPLMALGTAAYWALASGADEAPPRLRLVYSGASMLPRAVAGLAASGL 755
QY 286 CTAF--PYTIAVLYGI-----TDLDSILSSAGYIPFETMTQSLRSL 324
Db 756 ALLLWKPVTLVKAGAGAPRTRTTLTPFSGPPTSQADLDYVVPQI---YRHMQEERFGR- 810
QY 325 SFATVLSGGIVMAFFALNAVQETA--SRLTWSFARDNGLVFS---THLERIHPRW----- 375
Db 811 RLERTKSGPLTVAAYQLGVSYSAMVTALT-----LAFPLLLLHAERISLVFLLLF 863
```

QY 376 -----QVPV-----WSLFATWGIATCGCIFLGSSTAFNAL 406
Db 864 LQSFLLHLLAAGIPVTTGPGFTVPWQAVSAWALMATQTFYSTGHPVFPPI 915

RESULT 15

US-10-006-172A-102
; Sequence 102, Application US/10006172A

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan I.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: P2830PIC11
; CURRENT APPLICATION NUMBER: US/10/006,172A
; CURRENT FILING DATE: 2002-03-19
; PRIOR APPLICATION NUMBER: 60/098716
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098723
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098749
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098750
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098803
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/098821
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/098843
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/099536
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099596
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099598
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099602
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099642
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099741
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099754
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099763
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099792
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099808
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099812
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099815
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099816
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/100385
; PRIOR FILING DATE: 1998-09-15
; PRIOR APPLICATION NUMBER: 60/100388
; PRIOR FILING DATE: 1998-09-15

; PRIOR APPLICATION NUMBER: 60/100390
; PRIOR FILING DATE: 1998-09-15
; PRIOR APPLICATION NUMBER: 60/100584
; PRIOR FILING DATE: 1998-09-16
; PRIOR APPLICATION NUMBER: 60/100627
; PRIOR FILING DATE: 1998-09-16
; PRIOR APPLICATION NUMBER: 60/100661
; PRIOR FILING DATE: 1998-09-16
; PRIOR APPLICATION NUMBER: 60/100662
; PRIOR FILING DATE: 1998-09-16
; PRIOR APPLICATION NUMBER: 60/100664
; PRIOR FILING DATE: 1998-09-16
; PRIOR APPLICATION NUMBER: 60/100683
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: 60/100684
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: 60/100710
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: 60/100711
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: 60/100848
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 60/100849
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 60/100919
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: 60/100930
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: 60/101014
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 60/101068
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 60/101071
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 60/101279
; PRIOR FILING DATE: 1998-09-22
; PRIOR APPLICATION NUMBER: 60/101471
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101472
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101474
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101475
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101476
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101477
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101479
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101738
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/101741
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/101743
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/101915
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/101916
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/102207
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/102240
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/102307
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/102330
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/102331
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/102484
; PRIOR FILING DATE: 1998-09-30
; PRIOR APPLICATION NUMBER: 60/102487

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2002, 07:55:20 ; Search time 389.72 Seconds
(without alignments)
474.160 Million cell updates/sec

Title: US-09-882-694A-8
Perfect score: 2692
Sequence: 1 MDSRPSGYGEGGTRQTTKN.....QGPHELDGRVVGAEFQVGP 525

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 3502263 seqs, 351980561 residues

Total number of hits satisfying chosen parameters: 3502263

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents_AA_Main:*

1: /cgn2_6/ptodata/2/paa/pcrUS_COMB.pep:*

2: /cgn2_6/ptodata/2/paa/US06_COMB.pep:*

3: /cgn2_6/ptodata/2/paa/US07_COMB.pep:*

4: /cgn2_6/ptodata/2/paa/US080_COMB.pep:*

5: /cgn2_6/ptodata/2/paa/US081_COMB.pep:*

6: /cgn2_6/ptodata/2/paa/US082_COMB.pep:*

7: /cgn2_6/ptodata/2/paa/US083_COMB.pep:*

8: /cgn2_6/ptodata/2/paa/US084_COMB.pep:*

9: /cgn2_6/ptodata/2/paa/US085_COMB.pep:*

10: /cgn2_6/ptodata/2/paa/US086_COMB.pep:*

11: /cgn2_6/ptodata/2/paa/US087_COMB.pep:*

12: /cgn2_6/ptodata/2/paa/US088_COMB.pep:*

13: /cgn2_6/ptodata/2/paa/US089_COMB.pep:*

14: /cgn2_6/ptodata/2/paa/US090_COMB.pep:*

15: /cgn2_6/ptodata/2/paa/US091_COMB.pep:*

16: /cgn2_6/ptodata/2/paa/US092_COMB.pep:*

17: /cgn2_6/ptodata/2/paa/US093_COMB.pep:*

18: /cgn2_6/ptodata/2/paa/US094_COMB.pep:*

19: /cgn2_6/ptodata/2/paa/US095_COMB.pep:*

20: /cgn2_6/ptodata/2/paa/US096_COMB.pep:*

21: /cgn2_6/ptodata/2/paa/US097_COMB.pep:*

22: /cgn2_6/ptodata/2/paa/US098_COMB.pep:*

23: /cgn2_6/ptodata/2/paa/US099_COMB.pep:*

24: /cgn2_6/ptodata/2/paa/US100_COMB.pep:*

25: /cgn2_6/ptodata/2/paa/US101_COMB.pep:*

26: /cgn2_6/ptodata/2/paa/US60_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2692	100.0	525	17	US-09-351-224E-8
2	2692	100.0	525	20	US-09-677-488A-8
3	2692	100.0	525	20	US-09-677-682A-8
4	2692	100.0	525	20	US-09-677-682B-8
5	2692	100.0	525	22	US-09-882-694-8
6	2692	100.0	525	22	US-09-882-694A-8
7	2655	98.6	525	17	US-09-351-224-8

8	2655	98.6	525	17	US-09-351-8223-8	Sequence 8, Appli
9	2655	98.6	525	20	US-09-677-488-8	Sequence 8, Appli
10	2655	98.6	525	20	US-09-677-682-8	Sequence 8, Appli
11	561.5	20.9	585	16	US-09-248-796-20096	Sequence 20096, A
12	561.5	20.9	585	26	US-60-096-409-20096	Sequence 20096, A
13	552	20.5	503	1	PCT-US02-03987-15766	Sequence 15766, A
14	552	20.5	503	24	US-10-032-585-7822	Sequence 7822, Ap
15	552	20.5	503	24	US-10-072-851-15766	Sequence 15766, A
16	552	20.5	503	26	US-60-314-050-7822	Sequence 7822, Ap
17	544.5	20.2	450	16	US-09-248-796-20095	Sequence 20095, A
18	544.5	20.2	450	26	US-60-096-409-20095	Sequence 20095, A
19	519	19.3	411	18	US-09-417-507-33583	Sequence 33583, A
20	451	16.8	505	21	US-09-708-427-19713	Sequence 19713, A
21	451	16.8	517	21	US-09-708-427-19712	Sequence 19712, A
22	449.5	16.7	504	20	US-09-621-900A-72	Sequence 72, Appl
23	449.5	16.7	516	20	US-09-621-900A-71	Sequence 71, Appl
24	449.5	16.7	528	20	US-09-621-900A-70	Sequence 70, Appl
25	444.5	16.5	352	18	US-09-417-507-39634	Sequence 39634, A
26	400	14.9	438	21	US-09-708-427-19714	Sequence 19714, A
27	383.5	14.2	571	19	US-09-538-092-120	Sequence 120, App
28	363.5	13.5	533	16	US-09-248-796-20066	Sequence 20066, A
29	363.5	13.5	533	26	US-60-096-409-20066	Sequence 20066, A
30	349.5	13.0	381	16	US-09-248-796-20097	Sequence 20097, A
31	349.5	13.0	381	26	US-60-096-409-20097	Sequence 20097, A
32	343	12.7	426	26	US-60-324-109-17656	Sequence 17656, A
33	334	12.4	575	24	US-10-032-585-7921	Sequence 7921, Ap
34	333	12.4	268	20	US-09-675-784A-7032	Sequence 7032, Ap
35	332	12.3	475	16	US-09-248-796-20067	Sequence 20067, A
36	332	12.3	475	26	US-60-096-409-20067	Sequence 20067, A
37	292.5	10.9	202	18	US-09-417-507-24497	Sequence 24497, A
38	279.5	10.4	291	18	US-09-417-507-32921	Sequence 32921, A
39	268.5	10.0	182	18	US-09-417-507-34508	Sequence 34508, A
40	268.5	10.0	248	18	US-09-417-507-24006	Sequence 24006, A
41	268.5	10.0	283	26	US-60-324-109-25312	Sequence 25312, A
42	264.5	9.8	484	16	US-09-248-796-20357	Sequence 20357, A
43	264.5	9.8	484	26	US-60-096-409-20357	Sequence 20357, A
44	260	9.7	397	26	US-60-324-109-20988	Sequence 20988, A
45	255.5	9.5	287	18	US-09-417-507-37631	Sequence 37631, A

ALIGNMENTS

RESULT 1
US-09-351-224E-8
; Sequence 8, Application US/09351224E
; GENERAL INFORMATION:
; APPLICANT: Duwick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; TITLE OF INVENTION: Detoxification
; FILE REFERENCE: 5718-111
; CURRENT APPLICATION NUMBER: US/09/351,224E
; CURRENT FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 525
; TYPE: PRT
; ORGANISM: Exophiala spinifera
US-09-351-224E-8

Query Match 100.0%; Score 2692; DB 17; Length 525;
Best Local Similarity 100.0%; Pred. No. 3.8e-244;
Matches 525; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MDSRPSGYGEGGTRQTTKNTETAAGGASESLNPLEKKQFGTITIVSLAFVICSNWAG 60
|||||
Db 1 MDSRPSGYGEGGTRQTTKNTETAAGGASESLNPLEKKQFGTITIVSLAFVICSNWAG 60

```
QY 61 ISGSLQALLAGGPVTLTYGILISTLVYICIAFSLAELTSVYPTAGGOYHFASILAPKSI 120
|||||
Db 61 ISGSLQALLAGGPVTLTYGILISTLVYICIAFSLAELTSVYPTAGGOYHFASILAPKSI 120
|||||
QY 121 NRSISYVCGLVSLLSWIAIGSSVTMIPAAQOIPALIAAYSHTYSQDSWHVFLIYEGVALV 180
|||||
Db 121 NRSISYVCGLVSLLSWIAIGSSVTMIPAAQOIPALIAAYSHTYSQDSWHVFLIYEGVALV 180
|||||
QY 181 LFNLFALKRNPWVHEIGFGLTIALFVISFIALARSNPKAPNSQVWTAWSNYTGWSDGV 240
|||||
Db 181 LFNLFALKRNPWVHEIGFGLTIALFVISFIALARSNPKAPNSQVWTAWSNYTGWSDGV 240
|||||
QY 241 CFILGLSTSCFMFIGLDAAMHIAEECTDAARTVPKAVVSAIIIGFCTAFPTYIAVLYGIT 300
|||||
Db 241 CFILGLSTSCFMFIGLDAAMHIAEECTDAARTVPKAVVSAIIIGFCTAFPTYIAVLYGIT 300
|||||
QY 301 DLDSILSSAGYIPFETMTQSLRSLSFATVLSGGIVMAFFALNAVQETASRLTWSFARDN 360
|||||
Db 301 DLDSILSSAGYIPFETMTQSLRSLSFATVLSGGIVMAFFALNAVQETASRLTWSFARDN 360
|||||
QY 361 GLVFSTHLERIHPRWQVPVWSLFATWGILATCGCIFLGSSTAFNALVNSAVVLOQLSFLI 420
|||||
Db 361 GLVFSTHLERIHPRWQVPVWSLFATWGILATCGCIFLGSSTAFNALVNSAVVLOQLSFLI 420
|||||
QY 421 PIALLLYQKRDPKFLPSTRAFLVPRGIGFLVNLAVVFTSVTTVFFSFPLTVPTAASTMN 480
|||||
Db 421 PIALLLYQKRDPKFLPSTRAFLVPRGIGFLVNLAVVFTSVTTVFFSFPLTVPTAASTMN 480
|||||
QY 481 YTSAGIIGVALALGVLNVVHARKHYQGPHLELDGRVVGAEFQVGP 525
|||||
Db 481 YTSAGIIGVALALGVLNVVHARKHYQGPHLELDGRVVGAEFQVGP 525
|||||
RESULT 2
US-09-677-488A-8
; Sequence 8, Application US/09677488A
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; TITLE OF INVENTION: Detoxification
; FILE REFERENCE: 35718/204100
; CURRENT APPLICATION NUMBER: US/09/677,488A
; CURRENT FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 525
; TYPE: PRT
; ORGANISM: Exophiala spinifera
US-09-677-488A-8
Query Match 100.0%; Score 2692; DB 20; Length 525;
Best Local Similarity 100.0%; Pred. No. 3.8e-244;
Matches 525; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MDSRPSGYGKGGTROTTKNTETAAGGASESLNVPLEKKQFGTITIVSLAFVICNSWAG 60
|||||
Db 1 MDSRPSGYGKGGTROTTKNTETAAGGASESLNVPLEKKQFGTITIVSLAFVICNSWAG 60
|||||
QY 61 ISGSLQALLAGGPVTLTYGILISTLVYICIAFSLAELTSVYPTAGGOYHFASILAPKSI 120
|||||
Db 61 ISGSLQALLAGGPVTLTYGILISTLVYICIAFSLAELTSVYPTAGGOYHFASILAPKSI 120
|||||
QY 121 NRSISYVCGLVSLLSWIAIGSSVTMIPAAQOIPALIAAYSHTYSQDSWHVFLIYEGVALV 180
|||||
```

```
Db 121 NRSISYVCGLVSLLSWIAIGSSVTMIPAAQOIPALIAAYSHTYSQDSWHVFLIYEGVALV 180
QY 181 LFNLFALKRNPWVHEIGFGLTIALFVISFIALARSNPKAPNSQVWTAWSNYTGWSDGV 240
|||||
Db 181 LFNLFALKRNPWVHEIGFGLTIALFVISFIALARSNPKAPNSQVWTAWSNYTGWSDGV 240
|||||
QY 241 CFILGLSTSCFMFIGLDAAMHIAEECTDAARTVPKAVVSAIIIGFCTAFPTYIAVLYGIT 300
|||||
Db 241 CFILGLSTSCFMFIGLDAAMHIAEECTDAARTVPKAVVSAIIIGFCTAFPTYIAVLYGIT 300
|||||
QY 301 DLDSILSSAGYIPFETMTQSLRSLSFATVLSGGIVMAFFALNAVQETASRLTWSFARDN 360
|||||
Db 301 DLDSILSSAGYIPFETMTQSLRSLSFATVLSGGIVMAFFALNAVQETASRLTWSFARDN 360
|||||
QY 361 GLVFSTHLERIHPRWQVPVWSLFATWGILATCGCIFLGSSTAFNALVNSAVVLOQLSFLI 420
|||||
Db 361 GLVFSTHLERIHPRWQVPVWSLFATWGILATCGCIFLGSSTAFNALVNSAVVLOQLSFLI 420
|||||
QY 421 PIALLLYQKRDPKFLPSTRAFLVPRGIGFLVNLAVVFTSVTTVFFSFPLTVPTAASTMN 480
|||||
Db 421 PIALLLYQKRDPKFLPSTRAFLVPRGIGFLVNLAVVFTSVTTVFFSFPLTVPTAASTMN 480
|||||
QY 481 YTSAGIIGVALALGVLNVVHARKHYQGPHLELDGRVVGAEFQVGP 525
|||||
Db 481 YTSAGIIGVALALGVLNVVHARKHYQGPHLELDGRVVGAEFQVGP 525
|||||
RESULT 3
US-09-677-682A-8
; Sequence 8, Application US/09677682A
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; TITLE OF INVENTION: Detoxification
; FILE REFERENCE: 35718/204101
; CURRENT APPLICATION NUMBER: US/09/677,682A
; CURRENT FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 525
; TYPE: PRT
; ORGANISM: Exophiala spinifera
US-09-677-682A-8
Query Match 100.0%; Score 2692; DB 20; Length 525;
Best Local Similarity 100.0%; Pred. No. 3.8e-244;
Matches 525; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MDSRPSGYGKGGTROTTKNTETAAGGASESLNVPLEKKQFGTITIVSLAFVICNSWAG 60
|||||
Db 1 MDSRPSGYGKGGTROTTKNTETAAGGASESLNVPLEKKQFGTITIVSLAFVICNSWAG 60
|||||
QY 61 ISGSLQALLAGGPVTLTYGILISTLVYICIAFSLAELTSVYPTAGGOYHFASILAPKSI 120
|||||
Db 61 ISGSLQALLAGGPVTLTYGILISTLVYICIAFSLAELTSVYPTAGGOYHFASILAPKSI 120
|||||
QY 121 NRSISYVCGLVSLLSWIAIGSSVTMIPAAQOIPALIAAYSHTYSQDSWHVFLIYEGVALV 180
|||||
Db 121 NRSISYVCGLVSLLSWIAIGSSVTMIPAAQOIPALIAAYSHTYSQDSWHVFLIYEGVALV 180
|||||
QY 181 LFNLFALKRNPWVHEIGFGLTIALFVISFIALARSNPKAPNSQVWTAWSNYTGWSDGV 240
|||||
Db 181 LFNLFALKRNPWVHEIGFGLTIALFVISFIALARSNPKAPNSQVWTAWSNYTGWSDGV 240
|||||
QY 241 CFILGLSTSCFMFIGLDAAMHIAEECTDAARTVPKAVVSAIIIGFCTAFPTYIAVLYGIT 300
|||||
```

```

Db      241  CFIILGSTSCFMFIGLDAAMHAEECTDAARTVPKAVVSAIIIGFCTAFPTYIAVLYGIT 300
QY      301  DLDSILSSAGYIPFETMTQSLRSLSFATVLSGGIVMAFFALNAVQETASRLTWSFARDN 360
Db      301  DLDSILSSAGYIPFETMTQSLRSLSFATVLSGGIVMAFFALNAVQETASRLTWSFARDN 360
QY      361  GLVFSTHLERIHPRWQVPVWSLGFATWGLATCGCIFLGSSTAFNALVNSAVVLQQLSFLI 420
Db      361  GLVFSTHLERIHPRWQVPVWSLGFATWGLATCGCIFLGSSTAFNALVNSAVVLQQLSFLI 420
QY      421  PIALLLYQKRDPKFLPSTRAFLVPRGIGFLVNVLAHVFTSVTTVFFSFPLTVPTAASTMN 480
Db      421  PIALLLYQKRDPKFLPSTRAFLVPRGIGFLVNVLAHVFTSVTTVFFSFPLTVPTAASTMN 480
QY      481  YTSALIGVALALGVLNVVHARKHYQGPHLELDGRVVGAEFQVGP 525
Db      481  YTSALIGVALALGVLNVVHARKHYQGPHLELDGRVVGAEFQVGP 525
```

```

RESULT 4
US-09-677-682B-8
; Sequence 8, Application US/09677682B
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; FILE REFERENCE: 35718/204101
; CURRENT APPLICATION NUMBER: US/09/677,682B
; CURRENT FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 525
; TYPE: PRT
; ORGANISM: Exophiala spinifera
US-09-677-682B-8
```

```

Query Match      100.0%; Score 2692; DB 20; Length 525;
Best Local Similarity 100.0%; Pred. No. 3.8e-244;
Matches 525; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  MDSRPSGYGKGGTROTTKNTETAAGGASESLNVPLEKKQFGTITIVSLAFVICNSWAG 60
Db      1  MDSRPSGYGKGGTROTTKNTETAAGGASESLNVPLEKKQFGTITIVSLAFVICNSWAG 60
QY      61  ISGSLQALLAGGPVTLTYGILISTLVYICIAFSLAELTSVYPTAGGQYHFASILAPKSI 120
Db      61  ISGSLQALLAGGPVTLTYGILISTLVYICIAFSLAELTSVYPTAGGQYHFASILAPKSI 120
QY      121 NRSISYVCGLVSLLSWIAIGSSVTMIPAQQIPALIAAAYSHYTSQDSWHVFLIYEGVALV 180
Db      121 NRSISYVCGLVSLLSWIAIGSSVTMIPAQQIPALIAAAYSHYTSQDSWHVFLIYEGVALV 180
QY      181 LFLNLFALKRNPWWHEIGFGLTIALFVISFTAILARSNPKAPNSQVWTAWSNYTGWSDGV 240
Db      181 LFLNLFALKRNPWWHEIGFGLTIALFVISFTAILARSNPKAPNSQVWTAWSNYTGWSDGV 240
QY      241 CFILGLSTSCFMFIGLDAAMHAEECTDAARTVPKAVVSAIIIGFCTAFPTYIAVLYGIT 300
Db      241 CFILGLSTSCFMFIGLDAAMHAEECTDAARTVPKAVVSAIIIGFCTAFPTYIAVLYGIT 300
QY      301 DLDSILSSAGYIPFETMTQSLRSLSFATVLSGGIVMAFFALNAVQETASRLTWSFARDN 360
Db      301 DLDSILSSAGYIPFETMTQSLRSLSFATVLSGGIVMAFFALNAVQETASRLTWSFARDN 360
```

```

QY      361  GLVFSTHLERIHPRWQVPVWSLGFATWGLATCGCIFLGSSTAFNALVNSAVVLQQLSFLI 420
Db      361  GLVFSTHLERIHPRWQVPVWSLGFATWGLATCGCIFLGSSTAFNALVNSAVVLQQLSFLI 420
QY      421  PIALLLYQKRDPKFLPSTRAFLVPRGIGFLVNVLAHVFTSVTTVFFSFPLTVPTAASTMN 480
Db      421  PIALLLYQKRDPKFLPSTRAFLVPRGIGFLVNVLAHVFTSVTTVFFSFPLTVPTAASTMN 480
QY      481  YTSALIGVALALGVLNVVHARKHYQGPHLELDGRVVGAEFQVGP 525
Db      481  YTSALIGVALALGVLNVVHARKHYQGPHLELDGRVVGAEFQVGP 525

RESULT 5
US-09-882-694-8
; Sequence 8, Application US/09882694
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; FILE REFERENCE: 35718/208255
; CURRENT APPLICATION NUMBER: US/09/882,694
; CURRENT FILING DATE: 2001-06-12
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 525
; TYPE: PRT
; ORGANISM: Exophiala spinifera
US-09-882-694-8
```

```

Query Match      100.0%; Score 2692; DB 22; Length 525;
Best Local Similarity 100.0%; Pred. No. 3.8e-244;
Matches 525; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  MDSRPSGYGKGGTROTTKNTETAAGGASESLNVPLEKKQFGTITIVSLAFVICNSWAG 60
Db      1  MDSRPSGYGKGGTROTTKNTETAAGGASESLNVPLEKKQFGTITIVSLAFVICNSWAG 60
QY      61  ISGSLQALLAGGPVTLTYGILISTLVYICIAFSLAELTSVYPTAGGQYHFASILAPKSI 120
Db      61  ISGSLQALLAGGPVTLTYGILISTLVYICIAFSLAELTSVYPTAGGQYHFASILAPKSI 120
QY      121 NRSISYVCGLVSLLSWIAIGSSVTMIPAQQIPALIAAAYSHYTSQDSWHVFLIYEGVALV 180
Db      121 NRSISYVCGLVSLLSWIAIGSSVTMIPAQQIPALIAAAYSHYTSQDSWHVFLIYEGVALV 180
QY      181 LFLNLFALKRNPWWHEIGFGLTIALFVISFTAILARSNPKAPNSQVWTAWSNYTGWSDGV 240
Db      181 LFLNLFALKRNPWWHEIGFGLTIALFVISFTAILARSNPKAPNSQVWTAWSNYTGWSDGV 240
QY      241 CFILGLSTSCFMFIGLDAAMHAEECTDAARTVPKAVVSAIIIGFCTAFPTYIAVLYGIT 300
Db      241 CFILGLSTSCFMFIGLDAAMHAEECTDAARTVPKAVVSAIIIGFCTAFPTYIAVLYGIT 300
QY      301 DLDSILSSAGYIPFETMTQSLRSLSFATVLSGGIVMAFFALNAVQETASRLTWSFARDN 360
Db      301 DLDSILSSAGYIPFETMTQSLRSLSFATVLSGGIVMAFFALNAVQETASRLTWSFARDN 360
QY      361  GLVFSTHLERIHPRWQVPVWSLGFATWGLATCGCIFLGSSTAFNALVNSAVVLQQLSFLI 420
Db      361  GLVFSTHLERIHPRWQVPVWSLGFATWGLATCGCIFLGSSTAFNALVNSAVVLQQLSFLI 420
QY      421  PIALLLYQKRDPKFLPSTRAFLVPRGIGFLVNVLAHVFTSVTTVFFSFPLTVPTAASTMN 480
Db      421  PIALLLYQKRDPKFLPSTRAFLVPRGIGFLVNVLAHVFTSVTTVFFSFPLTVPTAASTMN 480
```


QY	481	YTS	I	G	V	A	L	G	V	L	N	W	V	H	A	R	K	H	Y	Q	G	P	H	L	E	D	G	R	V	V	G	A	E	F	F	V	G	P	525
Db	481	YTS	I	G	V	A	L	G	V	L	N	W	V	H	A	R	K	H	Y	Q	G	P	H	L	E	D	G	R	V	V	G	A	E	F	F	V	G	P	525

Query Match	100.0%;	Score 2692;	DB 22;	Length 525;
Best Local Similarity	100.0%;	pred. NO. 3.8e-244;		
Matches 525;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;

```

US-09-351-224-8
; Sequence 8, Application US/09351224
; GENERAL INFORMATION:
; APPLICANT: Duivick, Jon
; TITLE OF INVENTION: Compositions and Methods for Fummonisin Detoxification
; FILE REFERENCE: 5718-111
; CURRENT APPLICATION NUMBER: US/09/351,224
; CURRENT FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 525
; TYPE: PRT
; ORGANISM: Exophiala spinifera
; FEATURE:
; OTHER INFORMATION: permease
US-09-351-224-8

```

```
/ TYPE: PRT
; ORGANISM: Exophiala spinifera
; FEATURE:
; OTHER INFORMATION: permease
US-09-351-823-8

Query Match      98.6%; Score 2655; DB 17; Length 525;
Best Local Similarity 98.9%; Pred. No. 1.2e-240;
Matches 519; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 MDSRPSGYGKGGTROTTKNTETAAAGGASESLNVPLEKKQFGTITIVSLAFVICNSWAG 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 MDSRPSGYGLKGGTROTTKNTETAAAGGASESLNVPLEKKQFGTITIVSLAFVICNSWAG 60

Qy 61 ISGSLQALLAGGPVTLTYGILISTLVYICIAFSLAELTSVYPTAGGQYHFASILAPKSI 120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 ISGSLQALLAGGPVTLTYGILISTLVYICIAFSLAELTSVYPTAGGQYHFASILAPKSI 120

Qy 121 NRSISYVCGLVSLLSWIAIGSSVTMIPAQQIPALIAAAYSHYTSQDSWHVFLIYEGVALVV 180
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 NRSISYVCGLVSLLSWIAIGSSVTMIPAQQIPALIAAAYSHYTSQDSWHVFLIYEGVALVV 180

Qy 181 LLFNLFALKRNPWVHEIGFGLTIALFVISFIAILARSNPKAPNSQVWTAWSNYTGWSDGV 240
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 LLFNLFALKRNPWVHEIGFGLTIALFVISFIAILARSNPKAPNSQVWTAWSNYTGWSDGV 240

Qy 241 CFILGLSTSCFMFIGLDAAMHLEECTDAARTVPKAVVSAIIIGFCTAFPTYIAVLYGIT 300
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 241 CFILGLSTSCFMFIGLDAAMHLEECTDAARTVPKAVVSAIIIGFCTAFPTYIAVLYGIT 300

Qy 301 DLDSILSSAGYIPFETMTQSLRSLSFATVLSGCGIVMAFFALNAVQETASRLTWSFARDN 360
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 301 DLDSILSSAGYIPFETMRQSARIRLFATVLSGCGIVMAFFALNAVQETASRLTWSFARDN 360

Qy 361 GLVFSTHLERIHPRWQVPVWSLSFATWGILATCGCIFLGSSTAFNALVNSAVVLOQLSFLI 420
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 361 GLVFSTHLERIHPRWQVPVWSLSFATWGILATCGCIFLGSSTAFNALVNSAVVLOQLSFLI 420

Qy 421 PIALLLYQKRDPKFLPSTRAFLVLPARGIGFLVNLVAVVTSVTVFFSFPLTVPTAASTMN 480
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 421 PIALLLYQKRDPKFLPSTRAFLVLPARGIGFLVNLVAVVTSVTVFFSFPLTVPTAASTMN 480

Qy 481 YTSIIIGVALALGVLNWWVHARKHYQGPHLELDGRVVGAEFQVGP 525
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 481 YTSIIIGVALALGVLNWWVHARKHYQGPHLELDGRVVGAEFQVGP 525

RESULT 9
US-09-677-488-8
; Sequence 8, Application us/09677488
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jon
; TITLE OF INVENTION: Compositions and Methods for Fummonisin Detoxification
; FILE REFERENCE: 35718/204100 (5718-111A)
; CURRENT APPLICATION NUMBER: US/09/677,488
; CURRENT FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 525
; TYPE: PRT
; ORGANISM: Exophiala spinifera
; FEATURE:
; OTHER INFORMATION: permease, translation of fully spliced cDNA
US-09-677-488-8

Query Match      98.6%; Score 2655; DB 20; Length 525;
Best Local Similarity 98.9%; Pred. No. 1.2e-240;
Matches 519; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
```

```
Qy 1 MDSRPSGYGKGGTROTTKNTETAAAGGASESLNVPLEKKQFGTITIVSLAFVICNSWAG 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 MDSRPSGYGLKGGTROTTKNTETAAAGGASESLNVPLEKKQFGTITIVSLAFVICNSWAG 60

Qy 61 ISGSLQALLAGGPVTLTYGILISTLVYICIAFSLAELTSVYPTAGGQYHFASILAPKSI 120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 ISGSLQALLAGGPVTLTYGILISTLVYICIAFSLAELTSVYPTAGGQYHFASILAPKSI 120

Qy 121 NRSISYVCGLVSLLSWIAIGSSVTMIPAQQIPALIAAAYSHYTSQDSWHVFLIYEGVALVV 180
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 NRSISYVCGLVSLLSWIAIGSSVTMIPAQQIPALIAAAYSHYTSQDSWHVFLIYEGVALVV 180

Qy 181 LLFNLFALKRNPWVHEIGFGLTIALFVISFIAILARSNPKAPNSQVWTAWSNYTGWSDGV 240
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 LLFNLFALKRNPWVHEIGFGLTIALFVISFIAILARSNPKAPNSQVWTAWSNYTGWSDGV 240

Qy 241 CFILGLSTSCFMFIGLDAAMHLEECTDAARTVPKAVVSAIIIGFCTAFPTYIAVLYGIT 300
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 241 CFILGLSTSCFMFIGLDAAMHLEECTDAARTVPKAVVSAIIIGFCTAFPTYIAVLYGIT 300

Qy 301 DLDSILSSAGYIPFETMTQSLRSLSFATVLSGCGIVMAFFALNAVQETASRLTWSFARDN 360
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 301 DLDSILSSAGYIPFETMRQSARIRLFATVLSGCGIVMAFFALNAVQETASRLTWSFARDN 360

Qy 361 GLVFSTHLERIHPRWQVPVWSLSFATWGILATCGCIFLGSSTAFNALVNSAVVLOQLSFLI 420
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 361 GLVFSTHLERIHPRWQVPVWSLSFATWGILATCGCIFLGSSTAFNALVNSAVVLOQLSFLI 420

Qy 421 PIALLLYQKRDPKFLPSTRAFLVLPARGIGFLVNLVAVVTSVTVFFSFPLTVPTAASTMN 480
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 421 PIALLLYQKRDPKFLPSTRAFLVLPARGIGFLVNLVAVVTSVTVFFSFPLTVPTAASTMN 480

Qy 481 YTSIIIGVALALGVLNWWVHARKHYQGPHLELDGRVVGAEFQVGP 525
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 481 YTSIIIGVALALGVLNWWVHARKHYQGPHLELDGRVVGAEFQVGP 525

RESULT 10
US-09-677-682-8
; Sequence 8, Application US/09677682
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jon
; TITLE OF INVENTION: Compositions and Methods for Fummonisin Detoxification
; FILE REFERENCE: 35718/204101 (5718-111B)
; CURRENT APPLICATION NUMBER: US/09/677,682
; CURRENT FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 525
; TYPE: PRT
; ORGANISM: Exophiala spinifera
; FEATURE:
; OTHER INFORMATION: permease, translation of fully spliced cDNA
US-09-677-682-8

Query Match      98.6%; Score 2655; DB 20; Length 525;
Best Local Similarity 98.9%; Pred. No. 1.2e-240;
Matches 519; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
```

```
Qy 1 MDSRPSGYGKGGTROTTKNTETAAAGGASESLNVPLEKKQFGTITIVSLAFVICNSWAG 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 MDSRPSGYGLKGGTROTTKNTETAAAGGASESLNVPLEKKQFGTITIVSLAFVICNSWAG 60

Qy 61 ISGSLQALLAGGPVTLTYGILISTLVYICIAFSLAELTSVYPTAGGQYHFASILAPKSI 120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 ISGSLQALLAGGPVTLTYGILISTLVYICIAFSLAELTSVYPTAGGQYHFASILAPKSI 120

Qy 121 NRSISYVCGLVSLLSWIAIGSSVTMIPAQQIPALIAAAYSHYTSQDSWHVFLIYEGVALVV 180
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Db 121 NRSISYVCGLVSLLSWIAIGSSVTMIPAQQIPALIAAYSHYTSQDSWHVFLIYEGVALV 180
QY 181 LFPNLFALKRNPWWHEIGFGLTIALFVISFIAILARSNPKAPNSQVWTAWSNYTGWSDGV 240
Db 181 LFPNLFALKRNPWWHEIGFGLTIALFVISFIAILARSNPKAPNSQVWTAWSNYTGWSDGV 240
QY 241 CFTILGLSTSCFMFIGLDAAMHLAEECTDAARTVPKAVVSAILIIGFCTAFPTYTIAVLYGIT 300
Db 241 CFTILGLSTSCFMFIGLDAAMHLAEECTDAARTVPKAVVSAILIIGFCTAFPTYTIAVLYGIT 300
QY 301 DLDSILSSAGYIPFETMTQSLRSLSFATVLSGCGIVMAFFALNAVQETASRLTWSFARDN 360
Db 301 DLDSILSSAGYIPFETMQSARIRLFATVLSGCGIVMAFFALNAVQETASRLTWSFARDN 360
QY 361 GLVFSSTHLERIHPRWQVPVWSLFATWGILATCGCIFLGSSTAFNALVNSAVVLOQLSFLI 420
Db 361 GLVFSSTHLERIHPRWQVPVWSLFATWGILATCGCIFLGSSTAFNALVNSAVVLOQLSFLI 420
QY 421 PIALLLYQKRDPKFLPSTRAFLVPRGIGFLVNLAVVTSVTTVFESFPLTVPTAASPMN 480
Db 421 PIALLLYQKRDPKFLPSTRAFLVPRGIGFLVNLAVVTSVTTVFESFPLTVPTAASPMN 480
QY 481 YTSAILIGVALGVLNWMVHARKHYQGPHLELDGRVVGAEEFQVGP 525
Db 481 YTSAILIGVALGVLNWMVHARKHYQGPHLELDGRVVGAEEFQVGP 525

RESULT 11
US-09-248-796-20096
; Sequence 20096, Application US/09248796
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796
; CURRENT FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 28206
; SEQ ID NO 20096
; LENGTH: 585
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796-20096

Query Match 20.9%; Score 561.5; DB 16; Length 585;
Best Local Similarity 29.9%; Pred. No. 2e-43;
Matches 152; Conservative 106; Mismatches 210; Indels 41; Gaps 13;
QY 17 TTKNTETAAAGGASESLNVPLEKKQFGTTITVSLAFVICNSWAGISGSLQALLAGGPVT 76
Db 88 TSKDEDTKMS-----RNFSTLQIVTIGFGLTNSWLGLSTLILSISSLGPLL 134
QY 77 LLYGILISTLVYICIAFSLAELTSVYPTAGGQYHFASILAPKSINRSISYVCGLVSLLSW 136
Db 135 VVYGILIVASVSVCIATVLTGEMALAMPSAGGQYVWARVLAPKKYSSFLAYITGSISWGGA 194
QY 137 IAIGSSVTMIPAQQIPALIAAYSHYTSQDSWHVFLIYEGVALVVLLENLFALKRNPWWHE 196
Db 195 IFTTASMNLA VAYQVLGFWNMTHPDHVNQKWEVFIYINILNWILFFNWIH-RFLPMIGD 253
QY 197 IGFGLTIALFVISFIAILARSNPKAPNSQ-VWTAWSNYTGW-SDGVCFILGLSTSCFMFI 254
Db 254 SVFGISLTSYCIILITVLVCARGHYQDAKFVHVAVNNTGWPSKGIATVGLVNPAAWAFS 313
QY 314 CLDSVTHLSEETAHPERDVPRAVLSTVGIGFVTSFTFSIAMFFCIRNLEEIMNSATGFP 373
QY 314 FETMTQSLRSLSFATVLSGCGIVMAFFALNAVQETAS-----RLTWSFARDNGLVFS 367
Db 374 LDIFYQALGSTKVGA I--CLGSLITLVATGC---TLSCIIYQLRLLFSFLRDNGMPLSKY 428
QY 255 GLDAAMHLAEECTDAARTVPKAVVSAILIIGFCTAFPTYTIAVLYGITDLDSILSSA-GYIP 313
Db 314 CLDSVTHLSEETAHPERDVPRAVLSTVGIGFVTSFTFSIAMFFCIRNLEEIMNSATGFP 373
QY 254 SVFGISLTSYCIILITVLVCARGHYQDAKFVHVAVNNTGWPSKGIATVGLVNPAAWAFS 313
Db 314 CLDSVTHLSEETAHPERDVPRAVLSTVGIGFVTSFTFSIAMFFCIRNLEEIMNSATGFP 373
QY 314 FETMTQSLRSLSFATVLSGCGIVMAFFALNAVQETAS-----RLTWSFARDNGLVFS 367
Db 374 LDIFYQALGSTKVGA I--CLGSLITLVATGC---TLSCIIYQLRLLFSFLRDNGMPLSKY 428

QY 368 LERIHPRWQVPVWS-LFATWGILATCGCIFLGSSTAFNALVNSAVVLOQLSFLIPIALL 426
Db 429 LSIIDPKTGGPFYAHLPST--VIVSVSVLVFSDAALQATALACVPSFLLIAYLIPTICLL 486
QY 427 YQKRDPKFLPSTRAFLVPRGIGFLVNLAVVTSVTTVFESFPLTVPTAASPMNYSAI 486
Db 487 ARRRQIRHGP-----FWLGK-IGVFCNFVLLAWCIFA VVFFSFPANYPVTAEGMNYFCVVL 541
QY 487 GVALALGVLNWMVHARKH-----YQGPLH 510
Db 542 VVYIICMLGYWFWFPIKKYACKYNFRGGNL 570
RESULT 12
US-60-096-409-20096
; Sequence 20096, Application US/60096409A
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALB
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-016P
; CURRENT APPLICATION NUMBER: US/60/096,409A
; CURRENT FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28206
; SEQ ID NO 20096
; LENGTH: 585
; TYPE: PRT
; ORGANISM: Candida albicans
US-60-096-409-20096

Query Match 20.9%; Score 561.5; DB 26; Length 585;
Best Local Similarity 29.9%; Pred. No. 2e-43;
Matches 152; Conservative 106; Mismatches 210; Indels 41; Gaps 13;
QY 17 TTKNTETAAAGGASESLNVPLEKKQFGTTITVSLAFVICNSWAGISGSLQALLAGGPVT 76
Db 88 TSKDEDTKMS-----RNFSTLQIVTIGFGLTNSWLGLSTLILSISSLGPLL 134
QY 77 LLYGILISTLVYICIAFSLAELTSVYPTAGGQYHFASILAPKSINRSISYVCGLVSLLSW 136
Db 135 VVYGILIVASVSVCIATVLTGEMALAMPSAGGQYVWARVLAPKKYSSFLAYITGSISWGGA 194
QY 137 IAIGSSVTMIPAQQIPALIAAYSHYTSQDSWHVFLIYEGVALVVLLENLFALKRNPWWHE 196
Db 195 IFTTASMNLA VAYQVLGFWNMTHPDHVNQKWEVFIYINILNWILFFNWIH-RFLPMIGD 253
QY 197 IGFGLTIALFVISFIAILARSNPKAPNSQ-VWTAWSNYTGW-SDGVCFILGLSTSCFMFI 254
Db 254 SVFGISLTSYCIILITVLVCARGHYQDAKFVHVAVNNTGWPSKGIATVGLVNPAAWAFS 313
QY 255 GLDAAMHLAEECTDAARTVPKAVVSAILIIGFCTAFPTYTIAVLYGITDLDSILSSA-GYIP 313
Db 314 CLDSVTHLSEETAHPERDVPRAVLSTVGIGFVTSFTFSIAMFFCIRNLEEIMNSATGFP 373
QY 314 FETMTQSLRSLSFATVLSGCGIVMAFFALNAVQETAS-----RLTWSFARDNGLVFS 367
Db 374 LDIFYQALGSTKVGA I--CLGSLITLVATGC---TLSCIIYQLRLLFSFLRDNGMPLSKY 428
QY 368 LERIHPRWQVPVWS-LFATWGILATCGCIFLGSSTAFNALVNSAVVLOQLSFLIPIALL 426
Db 429 LSIIDPKTGGPFYAHLPST--VIVSVSVLVFSDAALQATALACVPSFLLIAYLIPTICLL 486
QY 427 YQKRDPKFLPSTRAFLVPRGIGFLVNLAVVTSVTTVFESFPLTVPTAASPMNYSAI 486
Db 487 ARRRQIRHGP-----FWLGK-IGVFCNFVLLAWCIFA VVFFSFPANYPVTAEGMNYFCVVL 541
QY 487 GVALALGVLNWMVHARKH-----YQGPLH 510
Db 542 VVYIICMLGYWFWFPIKKYACKYNFRGGNL 570

; CURRENT FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; NUMBER OF SEQ ID NOS: 15811
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15766
; LENGTH: 503
; TYPE: PRT
; ORGANISM: Candida albicans
US-10-072-851-15766

Query Match		20.5%;	Score 552;	DB 24;	Length 503;
Best Local Similarity		29.7%;	Pred. No. 1.3e-42;		
Matches 143;		Conservative 96;	Mismatches 215;	Indels 28;	Gaps 11;
QY	39	KKQFGTITIVSLAFVICNSWAGISGSLQALLAGGPVTLVYILISTLVYICIAFSLAEL 98			
Db	26	ERNFSLSTCAFQFTLLICSALAIGTFLSTVIGVGGSPVLIIFGFIVAITEFDLIICYSLAEL 85			
QY	99	TSVYPTAGGOYHFASILAPKSNRSISYVCGLVSLLSWIAIGSSVTMIPAAQIIPALIAAY 158			
Db	86	ASAYPHSSAQIHWTYCLASEKYKRSLSFLTGLSCAGWIFACFSSTYVASMFILALAQIY 145			
QY	159	SHYQSQDSWHVFLIYEGVALVLLFNLFALKRNPWVHEIGFGLTIALFVISF-----IA 212			
Db	146	HQDYVPKSFHYLLVYLAVFLSGYLVNVFLVKLLPLI-----TNISVAVINFGTFFIIT 199			
QY	213	ILARSNPKAPNSQVWTAMSNYTGW-SDGVCFILGLSTSCFMFIGLDAAMHLAEECTDAAR 271			
Db	200	LLVKS-PKQSAEFVFKNIINETGWSSNGVVFELGMLPSLACVTLFDGAVHLTDEIAQPER 258			
QY	272	TVPKAVVSAIIIGFCTAFPTYIAVLVYGITDLDLSILSSAGYIPF-ETMTQSLRSLSFATV- 329			
Db	259	NPLVMVISNTLSGVMAFFRAIVMFCVVNVNNSNPVCGEPIVQLMYDSFQSEALTITIG 318			
QY	330	LSCGGIVMAFFALNAVQETA-SRLTWSFARDNGLVFSHLEIRHPRWQVPVWSLFATWGI 388			
Db	319	VVC--LILTFVGSSYMYTSTSRLIWSFANSNGLPFSKYIGEVSNNLKSPVYALSFLTVL 376			
QY	389	LATCGCIFLGSSTAFNALVNSAVWLQQLSFLIPIALLLYQKRPKFLPSTR-----AFVL 443			
Db	377	CIITGLIMSGDGLNAVLTGSMVCINLSYLIPIACLLVKS---KFSTTHRFNERPYFCL 433			
QY	444	PRGIGFLVNVLAVVFTSVTVFFSFPLTVPTAASTMNYTSAIIGVALALGVNLNVVHARK 503			
Db	434	GK-FGLPMNIASVLWVCFINVWLNFPPLSYPTSDNMNACVVLGITCIIGIILWFVHGKS 492			
QY	504	HY 505			
Db	493	RY 494			

Search completed: April 27, 2002, 07:55:22
Job time: 583 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2002, 07:46:46 ; Search time 43.81 Seconds
(without alignments)
292.706 Million cell updates/sec

Title: US-09-882-694A-8
Perfect score: 2692
Sequence: 1 MDSRPSGYGKGGTRQTKN.....QGPHELDGRVVGAEFQVGP 525

Scoring table: BLOSUM62
Gapop 10.0 ; Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*

1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*

2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*

3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*

4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*

5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*

6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	164	6.1	622	2	US-08-132-990A-4
2	164	6.1	622	5	PCT-US92-09382-4
3	145	5.4	629	2	US-08-132-990A-8
4	145	5.4	629	5	PCT-US92-09382-8
5	137.5	5.1	511	2	US-09-073-362-1
6	137.5	5.1	511	2	US-09-243-920-1
7	129.5	4.8	515	2	US-09-073-362-3
8	129.5	4.8	515	2	US-09-243-920-3
9	126.5	4.7	503	4	US-09-068-195-24
10	116.5	4.3	453	2	US-08-132-990A-6
11	116.5	4.3	453	5	PCT-US92-09382-6
12	114	4.2	493	1	US-08-362-512A-4
13	114	4.2	493	4	US-08-964-939-4
14	109.5	4.1	513	4	US-09-097-889-15
15	107	4.0	438	2	US-08-677-049-9
16	107	4.0	506	4	US-09-066-047-7
17	104	3.9	367	2	US-08-132-990A-2
18	104	3.9	367	5	PCT-US92-09382-2
19	104	3.9	485	1	US-08-362-512A-2
20	104	3.9	485	4	US-08-964-939-2
21	98.5	3.7	418	4	US-09-030-267-5
22	97.5	3.6	459	4	US-09-097-889-22
23	97	3.6	434	2	US-08-677-049-7
24	97	3.6	472	4	US-09-354-129-2
25	97	3.6	1212	4	US-09-268-866-2
26	96.5	3.6	429	2	US-08-677-049-5
27	95	3.5	635	2	US-09-014-969-11

28	94.5	3.5	1353	4	US-09-398-193-99	Sequence 99, Appl
29	94	3.5	1165	1	US-08-240-357-2	Sequence 2, Appli
30	94	3.5	1294	2	US-08-819-288-3	Sequence 3, Appli
31	94	3.5	1294	4	US-09-400-348-3	Sequence 3, Appli
32	94	3.5	1321	1	US-08-261-822A-3	Sequence 3, Appli
33	94	3.5	1321	5	PCT-US95-07744A-3	Sequence 3, Appli
34	93.5	3.5	584	2	US-08-928-692-13	Sequence 13, Appl
35	93	3.5	1443	1	US-08-308-872B-2	Sequence 2, Appli
36	92.5	3.4	416	4	US-09-333-208-2	Sequence 2, Appli
37	92.5	3.4	416	4	US-09-333-254-2	Sequence 2, Appli
38	92.5	3.4	1305	4	US-08-864-785-3	Sequence 3, Appli
39	92.5	3.4	1353	3	US-08-894-173-2	Sequence 2, Appli
40	92.5	3.4	1353	4	US-09-398-193-2	Sequence 2, Appli
41	90.5	3.4	294	1	US-08-142-439A-4	Sequence 4, Appli
42	90.5	3.4	294	2	US-08-869-477-4	Sequence 4, Appli
43	90.5	3.4	350	1	US-08-118-270-41	Sequence 41, Appl
44	90.5	3.4	350	5	PCT-US93-08528-41	Sequence 41, Appl
45	90.5	3.4	1454	4	US-08-392-459-22	Sequence 22, Appl

ALIGNMENTS

RESULT 1

US-08-132-990A-4

; Sequence 4, Application US/08132990A

; Patent No. 5834589

; GENERAL INFORMATION:

; APPLICANT: MERUELO, DANIEL

; APPLICANT: YOSHIMOTO, TAKAYUKI

; TITLE OF INVENTION: Human Retrovirus Receptor and DNA Coding Therefor

; NUMBER OF SEQUENCES: 31

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Pennie & Edmonds

; STREET: 1155 Avenue of the Americas

; CITY: New York

; STATE: NY

; COUNTRY: USA

; ZIP: 10036

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.24

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/132,990A

; FILING DATE: 07-OCT-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/084,729

; FILING DATE: 29-JUN-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/US93/05569

; FILING DATE: 11-JUN-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/899,075

; FILING DATE: 11-JUN-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/806,178

; FILING DATE: 13-DEC-1991

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/627,950

; FILING DATE: 14-DEC-1990

; ATTORNEY/AGENT INFORMATION:

; NAME: Misrock, S. Leslie

; REGISTRATION NUMBER: 18,872

; REFERENCE/DOCKET NUMBER: 8105-004-999

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 790-9090

; TELEFAX: (212) 869-8864

; TELEX: 66441 PENNIE

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 622 amino acids

Db 314 YFCLDNNSPDPAFKHVGWEGAKYAVAVGSLCALS-----ASLLGSMFPM-----P 359

QY 351 RLWTSFARDNGLVFSTHLERIHPRWQVPVWSLFATWIGILATCGCIF-----LGSS 400

Db 360 RVIYAMAED-GLLFK-FLANVDRTKPTIATLASGAVAAVMAFLEDLKDLDLMSIGTL 417

QY 401 TAFNALVNSAVVLQOLSFLIPIALLLYQKRPK-----433

Db 418 LAYSLVAACVLVR-----YQEPQNLVYQMASTSDLPADQNELASTNDSQ 465

QY 434 --FLPSTRAFLV-----PRGI-----GFLVN-----VLAVVFTSVTTV 464

Db 466 LGFLPEAEMFSLKTLILSPKNMEPSKISGLIVNISTSLIAVLIITFCIVTVL 516

RESULT 5

US-09-073-362-1

; Sequence 1, Application US/09073362

; Patent No. 5942399

; GENERAL INFORMATION:

; APPLICANT: Hillman, Jennifer L.

; APPLICANT: Yue, Henry

; APPLICANT: Corley, Neil C.

; TITLE OF INVENTION: AMINO ACID PERMEASE HOMOLOG

; NUMBER OF SEQUENCES: 3

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Drive

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/073,362

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Cerrone, Michael C

; REGISTRATION NUMBER: 39,132

; REFERENCE/DOCKET NUMBER: PF-0514 US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 650-855-0555

; TELEFAX: 650-845-4166

; TELEX:

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 511 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; LIBRARY: ESOGTUT02

; CLONE: 2667831

US-09-073-362-1

Query Match 5.1%; Score 137.5; DB 2; Length 511;

Best Local Similarity 21.3%; Pred. No. 1.8e-05;

Matches 119; Conservative 90; Mismatches 216; Indels 133; Gaps 28;

QY 1 MDSRPSGYGKGGTRQTKNTKTETAAGGASESLNVPLEKKQFGTITIVSLAFVICNSWAG 60

Db 15 VETSPIGDCASPGPEQVKLKKREISLNGVCLVIG-----NMIGSGIFVSPKGVLIYS--- 66

QY 61 ISGSLQALLAGGPVTLLYGILISTLVYICIAFSLAELTSVYPTAGGQYHPASILAPKSI 120

Db 67 ASEGLSLVIAVGGGLFSVFG-----ALCYAELGTTIKKSGASY----- 104

QY 121 NRSISYVC-GLVSLLSWIAIGSSVTMI-PAQQ-----IPALIAAYSHYTS 163

Db 105 ----AYILEAFGGFLAFIRLWTSLLIIEPTSOAILAITFANYMVQPLFPSCFAPYAASR- 159

QY 164 ODSWHVELIYEGVALVLLFNLFALKRNPWVHEIGFGLTIALFVISFIALARSNPKAPN 223

Db 160 -----LLAAACICLLTFINCAYVVKWGLVQDI---FTYAK-VLALIAVIAGIVRLGQ 208

QY 224 SOVWTAMSNYTGWSGDGVCFILG-----LSYSCPMFIFGLDAAMHLAEECTDAARTVPKAV- 277

Db 209 G----ASTHFENSFEGSSFAVGDIALALYSALFSYSGWDTLNYVYTEIKNPERNPLSIG 264

QY 278 VSAIIIGFCTAPPYTIAVLYGITDLSILSS-AGYIPFETMTQSLRS--LSFATVLSGG 334

Db 265 ISMPIV--TIIYILTNAVYYTVLDMRDILASDAVAVTFADQIFGIFENWIIPLSVALSC-- 320

QY 335 IVMAFFALNAVQETASRLTWSFARONGL---VFSTHLERIHPRWQVPVWSLFATWIGILAT 391

Db 321 ---FGGLNASIVAASRLFEVGSREGHLPDALCMIHVERF-----TPVPSLLFN-GIMA- 369

QY 392 CGCIFLGSSTAFNALVNSAVVLQQLSF-----LIPIALLLYQKRPDKFLPSTRAFVLP 444

Db 370 --LIYLCVEDIEFQ-----LINYSFSYWFVGLSIGQLYLRWKEPD-----RP 411

QY 445 RGIGFLVNVLAVVFTSVTTVFFSFPLTVPTAASTMNYTSAIIGVALALGVLNW--VVHA 501

Db 412 RPL-----KLSVFFFPVFCCLCTIFLVAVPLYSDTIN--SLIGIAIALSGLPFYFLIIRV 463

QY 502 RKHYQGPHELDGRVVG 519

Db 464 PEKRPPLYLR--RIVGS 478

RESULT 6

US-09-243-920-1

; Sequence 1, Application US/09243920

; Patent No. 5981242

; GENERAL INFORMATION:

; APPLICANT: Hillman, Jennifer L.

; APPLICANT: Yue, Henry

; APPLICANT: Corley, Neil C.

; TITLE OF INVENTION: AMINO ACID PERMEASE HOMOLOG

; NUMBER OF SEQUENCES: 3

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Drive

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/243,920

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 09/073,362

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Cerrone, Michael C

; REGISTRATION NUMBER: 39,132

; REFERENCE/DOCKET NUMBER: PF-0514 US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 650-855-0555

; TELEFAX: 650-845-4166

QY 340 ---FALNAVQETASRLTWSFARDNGLVSTHLEIRHPRW-----QVPV-----WSLF 383
Db 287 ALMIARFVGMGEVSS---WVVGPSRGM-FAAAQRLPKFLRKTNTHIEVPVPLVMIQGIIV 342
QY 384 ATWGILATCGCIFLGSSTAFNALVNSAVVLQ---QLSFLIPIALLLYQKRDPKFLPSTRA 440
Db 343 TLWGAULTFG--GGGNLSFLVAISLTVVIVLVGYLLFFVIVFVLIYKKQNLK-----RT 395
QY 441 FVLP-----RGIGFLVNLAVVFTSVTTVFFSEFPLTVPTAASTMNYT 482
Db 396 YNVPKTIKGTIIAGIGFLLSIFA-----LFISF---VPPASIAKNET 435
RESULT 10
US-08-132-990A-6
; Sequence 6, Application US/08132990A
; Patent No. 5834589
; GENERAL INFORMATION:
; APPLICANT: MERUELO, DANIEL
; APPLICANT: YOSHIMOTO, TAKAYUKI
; TITLE OF INVENTION: Human Retrovirus Receptor and DNA Coding Therefor
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/132,990A
; FILING DATE: 07-OCT-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/084,729
; FILING DATE: 29-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/05569
; FILING DATE: 11-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/899,075
; FILING DATE: 11-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/806,178
; FILING DATE: 13-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/627,950
; FILING DATE: 14-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 8105-004-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66441 PENNIE
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 453 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-132-990A-6

Query Match 4.3%; Score 116.5; DB 2; Length 453;
Best Local Similarity 21.9%; Pred. No. 0.002;
Matches 71; Conservative 55; Mismatches 119; Indels 79; Gaps 17;

QY 211 IAILARSNPKAPNSQVWTAWS-----NYTGWSDGVCFILGLSTSCFMFIGLDAAMHLAEEC 266
Db 23 ISASAREPPSENGTSIYGAGGFMPYGTG-----TLAGAAATCFYAFVGFDCIATTGEEV 76
QY 267 TDAARTVPKAVVSAIIIGFCTAF-----PYTIAVLYGITDLDLSILSSA-----GYIPFET 316
Db 77 RNPQAIPIGIVTSLVLCFMAYFGVSAALTLMPYLLDEKSPLPVAFEYVVRWGPACYVV 136
QY 317 MTQSLRSLSFATVLSGGIVMAFFALNAVOETASRLTWSFARDNGLVSTHLEIRHPRWQ 376
Db 137 AAGSLCALSTS-----LLGSIFPM-----PRVIYAMAED-GLLFKC-LAQINSKTK 180
QY 377 VPVWSLPATWGILATCGCIFLGSSTAFNALVN--SAVVLOQLSFLIPIALL-YQK---- 429
Db 181 TPVIATLSSGAVAAVMAFLF-----DLKALVDMMSIGTLMAYSLVAACVLILRYQPGLCY 235
QY 430 RDPKFLP-----STRAFVLPRGIGFLVNLAVVFTSVTTVFFSFPLTVPTA 475
Db 236 EQPKYTPKETLESCTNATLKSESOVTMLQGGF-----SLRTLTF--SPSALPTR 283
QY 476 ASTMNYTSATIG---VALALGVL 495
Db 284 QSA-SLVSVFLVGLAFLILGLSIL 306
RESULT 11
PCT-US92-09382-6
; Sequence 6, Application PC/TUS9209382
; GENERAL INFORMATION:
; APPLICANT: MERUELO, DANIEL
; APPLICANT: YOSHIMOTO, TAKAYUKI
; TITLE OF INVENTION: Human Retrovirus Receptor and DNA Coding
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/09382
; FILING DATE: 19921213
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Livnat, Shmuel
; REGISTRATION NUMBER: 33,949
; REFERENCE/DOCKET NUMBER: MERUELO-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 628-5197
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 453 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US92-09382-6

Query Match 4.3%; Score 116.5; DB 5; Length 453;
Best Local Similarity 21.9%; Pred. No. 0.002;
Matches 71; Conservative 55; Mismatches 119; Indels 79; Gaps 17;

QY 211 IAILARSNPKAPNSQVWTAWS-----NYTGWSDGVCFILGLSTSCFMFIGLDAAMHLAEEC 266
Db 23 ISASAREPPSENGTSIYGAGGFMPYGTG-----TLAGAAATCFYAFVGFDCIATTGEEV 76

Db 23 ISASAREPPSENGTSIYGAGGMPYGF--TLGAATCFYAFVGFDCIATGEEV 76
QY 267 TDAARTVPKAVSAIIGFCTAF---PYTIAVLYGITDLSILSSA-----GYIPFET 316
Db 77 RNPQKAIPIGIVTSLVCFMAYFGVSAALTLMPYLLDEKSPLPVAFYVVRWGPAYV 136
QY 317 MTQSLRSLSFATVSLCGGIVMAFFALNAVQETASRLTWSFARDNGLVFSTHLERHPRWQ 376
Db 137 AAGSLCALSTS-----LLGSIFFM-----PRVIYAMAED-GLLEKFC-LAQINSKTK 180
QY 377 VPVWSLFATWGILATCGCIFLGSSTAFNALVN--SAVVLOQLSFLIPIALL-YQK--- 429
Db 181 TPVIATLSSGAVAAVMAFLF-----DLKALVDMMSIGTLMAYSLVAACVLILRYQPGLCY 235
QY 430 RDPKFLP-----STRAFLVPRGIGFLVNVLAUVFTSVTTVFFSFPLTVPTA 475
Db 236 EQPKYTPEKETLESCNTNATLKSESVTMLQGGF-----SLRTLFL-SPSALPTR 283
QY 476 ASTMNYTSAIIG---VALALGVL 495
Db 284 QSA-SLVSFVLVGLAFLILGLSIL 306

RESULT 12

US-08-362-512A-4
; Sequence 4, Application US/08362512A
; Patent No. 5719043
; GENERAL INFORMATION:
; APPLICANT: FROMMER, Wolf-Bernd
; TITLE OF INVENTION: DNA SEQUENCES FOR AN AMINO ACID
; TITLE OF INVENTION: TRANSPORTER, PLASMIDS, BACTERIA, YEASTS AND PLANTS
; TITLE OF INVENTION: CONTAINING A TRANSPORTER AND THEIR USE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen
; STREET: 1180 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: US
; ZIP: 10036-8403
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/362,512A
; FILING DATE: 05-JAN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/007,636
; FILING DATE: 21-JAN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Meilman, Edward A.
; REGISTRATION NUMBER: 24,735
; REFERENCE/DOCKET NUMBER: P/951-107
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 382-0700
; TELEFAX: (212) 382-0888
; TELEX: 236925
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 493 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

Query Match 4.2%; Score 114; DB 1; Length 493;
Best Local Similarity 21.9%; Pred. No. 0.0041;
Matches 111; Conservative 72; Mismatches 149; Indels 174; Gaps 32;

QY 40 KQFGT-----ITIVSLAFVICONSWAGISGSLQALLAGGPVTLTYGILISTLVYICI 91
Db 45 KRTGTVMTASAHIIITAVIGSVLSLAWA-----IAQLGWIAGPAVMLLFSLV--TL----- 93
QY 92 AFSLAELTSSVYPT--AGQYHFASILAPKSNRSISY-VCGLVSLLS--WIAIGSSVTM 145
Db 94 -YSSTLLSDCYRTGDVSGKRNITYMDAVRSILGGFKFKICGLIQYLNLFGLAIGYTI-- 150
QY 146 IPAQOIPALIAAYSHTY----SQDSWHVFLIYEGVALVLLFNLFALKRNPVWHEIGFGL 201
Db 151 --AASISNMAIKRSCNCFHKSGGKDPCH-----MSSNPYM--IVFG- 186
QY 202 TIALFVISFIAILARSNPKAPN-SQVWTANSNYTGWSDGVCFILGLSTSCFMFICLDAAM 260
Db 187 -VAEILLS-----QVPDFDQIW-----WISIVAAMVSFTYSA---IGL--AL 222
QY 261 HLAEECTDAARTVPKAVVSAIIGFCT-----AFPTYTIAVLYGITDLSI 305
Db 223 GIVQV---AANGVFKGSLTGISIGTVTQTKIWRFTFOALGDIAFAYSYSVVL-IEIQDTV 278
QY 306 LSSAGYIPFE--TMTQSLR-SLSFATV--LSCGGIVMAFFALNAVQETASRLTWSFARDN 360
Db 279 RSP-----PAESKTMKKATKISIAVTTTFYMLCGSMGYAAG-----DAAPGNLLTGFGFYN 330
QY 361 GLVFSTHLERHPRWQVPVWSLFPATWGILATCGCIFLGSSTAFNALVNSAVVLOQLS--- 417
Db 331 -----PFWLL-----DIANAIVVHLVGAYQ 351
QY 418 -FLIPIALLLYQK-----RDPKFLPSTRAFLVPRGIGF----LVNVLAUVFTSVTTVFFS 467
Db 352 VFAOPIFAFIEKSAERYPDNDFLSKEFEIRIP---GFKSPYKVNVRMVRYS-----G 402
QY 468 FPLTVPTAASTMNYTSAIIGVALALG 493
Db 403 FVTTTIVISMLMPFFNDVVVGILGALG 428

RESULT 13

US-08-964-939-4
; Sequence 4, Application US/08964939
; Patent No. 6245970
; GENERAL INFORMATION:
; APPLICANT: FROMMER, Wolf-Bernd
; TITLE OF INVENTION: DNA SEQUENCES FOR AN AMINO ACID
; TITLE OF INVENTION: TRANSPORTER, PLASMIDS, BACTERIA, YEASTS AND PLANTS
; TITLE OF INVENTION: CONTAINING A TRANSPORTER AND THEIR USE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen
; STREET: 1180 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: US
; ZIP: 10036-8403
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/964,939
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,512
; FILING DATE: 05-JAN-1995
; APPLICATION NUMBER: US 08/007,636
; FILING DATE: 21-JAN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Meilman, Edward A.
; REGISTRATION NUMBER: 24,735
; REFERENCE/DOCKET NUMBER: P/951-107
; TELECOMMUNICATION INFORMATION:

; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: PURIFIED MAMMALIAN NUCLEOBASE PERMEASES;
; TITLE OF INVENTION: NUCLEIC ACIDS; ANTIBODIES
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DNAX Research Institute
; STREET: 901 California Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/677,049
; FILING DATE: 03-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/000,788
; FILING DATE: 03-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Ching, Edwin P.
; REGISTRATION NUMBER: 34,090
; REFERENCE/DOCKET NUMBER: DX0511
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-852-9196
; TELEFAX: 415-496-1200
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 438 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Region
; LOCATION: 115..144
; OTHER INFORMATION: /note= "Encompasses TM 4 of Figure
; OTHER INFORMATION: 4"
; FEATURE:
; NAME/KEY: Region
; LOCATION: 323..357
; OTHER INFORMATION: /note= "Encompasses TM 9 of Figure
; OTHER INFORMATION: 4"
; FEATURE:
; NAME/KEY: Region
; LOCATION: 359..386
; OTHER INFORMATION: /note= "Encompasses TM 10 of Figure
; OTHER INFORMATION: 4"
;
US-08-677-049-9

Query Match 4.0%; Score 107; DB 2; Length 438;
Best Local Similarity 19.8%; Pred. No. 0.018;
Matches 102; Conservative 78; Mismatches 172; Indels 164; Gaps 25;

QY 42 FGTITIVSLAFVICSNWAGISGLQALLAGGPVTL--YGILISTLVYICIAFSLAEL 98
|| : : | : : : | : | : | : : : : : : : : : : : : : : :
Db 5 FGKTLSLGIQHVL---AMYAGAIIVPLIVGKAMGLTVEQLTYLVSIDIFMCGVATLLQV 60

QY 99 -----TSVYP--TAGGQYH----FASILAPKSNRSISYVCG-LVSL 133
| : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 61 WSNRFFGICLPVVLGCTFTAVSPMAIGSEYGVSTVYGSIIASGILVILISFFFGKLVSF 120

QY 134 LSWIATGSSVTMIPAQQIPALIAAYSHYTSQDSWHVFLIYEGVALVVLLENLFALKRNPW 193
: || || : | : | : | : | : | : | : | : | : | : | : | :
Db 121 FPPVVTGSVVTII-----GITLMPVAMNNMAGGEG-- 150

QY 194 VHEIGFG----LTIALFVISFIALLARSNKAPNSQVWTAWNSYTGWSDGVCFILGLSTS 249
|| | : | | : | : | : | : | : | : | : | : | : | : | :

Db 151 --SADFGDLSNLALAFVLSIIVLLYRFT-----KGFIKSVSILIGLIG 193
QY 250 CFM--FIG-----LDAAM-----HLAEECTDAARTVVPKAVVSAIIIGFCTAPPVTI 293
| : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 194 TFIAYFMGKVQFDNVSDAAVQMIQPFYFGAPSFHAAPITMSIVAIVSLVESTGVYFAL 253
QY 294 AVLYG--ITDLDLSILSSAGYIPFETMTQSLRSLSPATVLSGCGIYMAFFALNAVQETASR 351
| : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 254 GDLTNRRLTEID---LSKGY-----RAEGLAVLL--GGI-----FNAFPYTA-- 290
QY 352 LTWSFARDNGLVFSTHLERIHPRWQVPVWSLSLPATWGILATCGCIFLG-----SSTAFNAL 406
| : : : | : | :
Db 291 ----FSQNVGLVQLTGIKN-----AVIVVTGVILMAFGLFPKIAAFTTI 331
QY 407 VNSAV----VLOQLSFLIPIALLLYQKRDPKFLPSTRAFLVPRGIGFLVNVLAVVFTSVT 462
: ||| : : | : : | : | : : : : : : : : : : : : : :
Db 332 IPSAVLGGAMVAMFGMVIAYGIKMLSRID--PAQENLLIVACSVG-----LGLGVTVVP 384
QY 463 TVFFSFPLTVPTAASTMNYTSAIIGVALALGVLNVV 498
: | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 385 DIFKQLP-----SALTLLTTNGIVAGSFTAVVLNIV 415

Search completed: April 27, 2002, 07:46:48
Job time: 6705 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2002, 07:45:33 ; Search time 104.96 Seconds
(without alignments)
555.581 Million cell updates/sec

Title: US-09-882-694A-8
Perfect score: 2692
Sequence: 1 MDSRPSGYGKGGTQTTKN.....QGPHELDGRVVGAEFQVGP 525

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802.*
1: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2655	98.6	525	22 AAY72636	Exophiala spinifera
2	451	16.8	505	21 AAG31987	Arabidopsis thalia
3	451	16.8	517	21 AAG31986	Arabidopsis thalia
4	449.5	16.7	504	21 AAG20282	Arabidopsis thalia
5	449.5	16.7	504	21 AAG49885	Arabidopsis thalia
6	449.5	16.7	516	21 AAG20281	Arabidopsis thalia
7	449.5	16.7	516	21 AAG49884	Arabidopsis thalia
8	449.5	16.7	528	21 AAG20280	Arabidopsis thalia
9	400	14.9	438	21 AAG31988	Arabidopsis thalia
10	398.5	14.8	437	21 AAG49886	Arabidopsis thalia
11	326.5	12.1	439	21 AAG49911	Arabidopsis thalia

12	326.5	12.1	521	21	AAG49910	Arabidopsis thalia
13	305	11.3	311	21	AAG49912	Arabidopsis thalia
14	216	8.0	482	22	AAU37348	Staphylococcus aur
15	204.5	7.6	453	22	AAG90272	C glutamicum prote
16	197	7.3	469	22	AAU52137	Propionibacterium
17	192	7.1	412	22	AAG81637	S. epidermidis ope
18	190	7.1	456	22	AAU33436	Enterococcus faeca
19	190	7.1	463	22	AAU33436	Enterococcus faeca
20	182	6.8	508	22	AAU40521	Propionibacterium
21	179.5	6.7	475	22	AAU33559	Pseudomonas aerugi
22	176	6.5	530	22	AAG63803	Amino acid sequenc
23	175	6.5	499	22	ABB71209	Drosophila melanog
24	173	6.4	508	22	AAU39612	Propionibacterium
25	171	6.4	466	22	AAU38455	Salmonella typhi c
26	169.5	6.3	462	22	AAG98875	E. coli growth and
27	169	6.3	390	22	AAG82164	S. epidermidis ope
28	167	6.2	544	22	AAU40519	Propionibacterium
29	164.5	6.1	489	22	AAU34622	E. coli cellular p
30	164.5	6.1	489	22	AAG64107	Escherichia coli l
31	164	6.1	523	22	AAG63804	Amino acid sequenc
32	164	6.1	622	13	AAR28335	ERR receptor. Mus
33	164	6.1	622	20	AAW67474	Murine ecotropic r
34	163.5	6.1	804	22	ABB69136	Drosophila melanog
35	162	6.0	466	22	AAU34672	E. coli cellular p
36	162	6.0	540	22	AAU57054	Propionibacterium
37	161	6.0	466	14	AAR39523	Sequence of gamma-
38	160	5.9	880	22	ABB63131	Drosophila melanog
39	159	5.9	489	22	AAU38339	Salmonella typhi c
40	159	5.9	510	21	AAG40406	Arabidopsis thalia
41	159	5.9	585	21	AAG40405	Arabidopsis thalia
42	159	5.9	590	21	AAG40404	Arabidopsis thalia
43	158.5	5.9	455	22	AAU64449	Propionibacterium
44	157.5	5.9	533	22	AAB72393	Rat L-type amino a
45	157.5	5.9	604	22	ABB65477	Drosophila melanog

ALIGNMENTS

RESULT 1

AAAY72636
ID AAY72636 standard; Protein; 525 AA.

XX

AC AAY72636;

XX

DT 02-MAY-2001 (first entry)

XX

DE Exophiala spinifera permease, a funonisin degradative enzyme.

XX

XW Fumonisin; degradative enzyme; transport enzyme; flavin monooxygenase;

KW aldehyde dehydrogenase; permease; p-glycoprotein; fumonisin esterase;

KW detoxification; mycotoxin; animal feed; human feed; silage;

KW transgenic plant; transgenic animal; microbial spray.

XX

OS Exophiala spinifera.

XX

PN WO200105980-A1.

XX

PD 25-JAN-2001.

XX

PF 14-JUL-1999; 99WO-US15824.

XX

PR 14-JUL-1999; 99WO-US15824.

XX

PA (PION-) PIONEER HI-BRED INT INC.

PA (CURA-) CURAGEN CORP.

XX

PI Duvick JP, Maddox J, Gilliam J, Folkerts O, Crasta OR;

XX

DR WPI; 2001-147345/15.

DR N-PSDB; AAD02693.

XX

PT Novel polynucleotides encoding Exophiala degradative or transport

PT enzyme which is useful for detoxifying fumonisin or structurally
 PT related mycotoxin during processing of grain for human or animal food
 PT consumption -
 XX
 PS Claim 1c: Page 71-73; 90pp; English.
 XX
 CC The patent discloses novel polynucleotides encoding *Exophiala spinifera*
 CC fumonisin degradative or transase enzymes such as flavin monooxygenase,
 CC aldehyde dehydrogenase, permease and p-glycoprotein. A fumonisin esterase,
 CC enzyme with at least one fumonisin degradative enzyme is useful for
 CC detoxifying fumonisin or a structurally related mycotoxin present in
 CC harvested grain, on application to a plant or to harvested grain
 CC during processing, or to processed grain that is to be used as animal
 CC or human feed, or as a silage. Nucleotide fragments of the present
 CC invention are useful as probes and primers. They can be introduced
 CC into microorganisms that multiply on plants to deliver enzymes to
 CC potential target crops. The genes encoding the degrading enzymes are
 CC introduced via a vector into a microbial host and the transformed host
 CC is supplied to the environment, plants or animals for reducing the
 CC pathogenicity of a fungus producing fumonisin. The genes of the
 CC invention are fermented in a bacterial host and the resulting bacteria
 CC is processed and used as a microbial spray. The nucleotide sequences
 CC can be used alone or in combination to engineer microbes or other
 CC organisms to metabolise fumonisin and resist its toxic effects.
 CC The present protein sequence is permease, a fumonisin degradative
 CC enzyme from *Exophiala spinifera*.
 XX
 SQ Sequence 525 AA;

AA	1987	standard; Protein; 505	AA.
AA	1987;		
DT	17-OCT-2000	(first entry)	
XX	Arabidopsis	thaliana protein fragment	SEQ ID NO: 38507.
DE	Protein identification; signal transduction pathway; metabolic pathway;		
XX	hybridisation assay; genetic mapping; gene expression control; promoter;		
KW	termination sequence.		
KW	Arabidopsis thaliana.		
XX	EP1033405-A2.		
OS	06-SEP-2000.		
XX	25-FEB-2000; 2000EP-0301439.		
PN	25-FEB-1999;	99US-0121825.	
XX	05-MAR-1999;	99US-0123180.	
PR	09-MAR-1999;	99US-0123548.	
PR	23-MAR-1999;	99US-0125788.	
PR	25-MAR-1999;	99US-0126264.	
PR	29-MAR-1999;	99US-0126785.	
PR	01-APR-1999;	99US-0127462.	
PR	06-APR-1999;	99US-0128234.	
PR	08-APR-1999;	99US-0128714.	
PR	16-APR-1999;	99US-0129845.	
PR	19-APR-1999;	99US-0130077.	
PR	21-APR-1999;	99US-0130449.	
PR	23-APR-1999;	99US-0130510.	
PR	23-APR-1999;	99US-0130891.	
PR	28-APR-1999;	99US-0131449.	
PR	30-APR-1999;	99US-0132048.	
PR	30-APR-1999;	99US-0132407.	
PR	04-MAY-1999;	99US-0132484.	
PR	05-MAY-1999;	99US-0132485.	
PR	06-MAY-1999;	99US-0132486.	
PR	07-MAY-1999;	99US-0132487.	
PR	11-MAY-1999;	99US-0132863.	
PR	11-MAY-1999;	99US-0134256.	
PR	14-MAY-1999;	99US-0134218.	
PR	14-MAY-1999;	99US-0134219.	
PR	14-MAY-1999;	99US-0134221.	
PR	14-MAY-1999;	99US-0134370.	
PR	18-MAY-1999;	99US-0134768.	
PR	19-MAY-1999;	99US-0134941.	
PR	20-MAY-1999;	99US-0135124.	
PR	21-MAY-1999;	99US-0135353.	
PR	24-MAY-1999;	99US-0135629.	
PR	25-MAY-1999;	99US-0136021.	
PR	27-MAY-1999;	99US-0136392.	
PR	28-MAY-1999;	99US-0136782.	
PR	01-JUN-1999;	99US-0137222.	
PR	03-JUN-1999;	99US-0137528.	
PR	04-JUN-1999;	99US-0137502.	
PR	07-JUN-1999;	99US-0137724.	
PR	08-JUN-1999;	99US-0138094.	
PR	10-JUN-1999;	99US-0138540.	
PR	10-JUN-1999;	99US-0138847.	
PR	14-JUN-1999;	99US-0139119.	
PR	16-JUN-1999;	99US-0139452.	
PR	16-JUN-1999;	99US-0139453.	
PR	17-JUN-1999;	99US-0139492.	
PR	18-JUN-1999;	99US-0139454.	
PR	18-JUN-1999;	99US-0139455.	
PR	18-JUN-1999;	99US-0139456.	
PR	18-JUN-1999;	99US-0139457.	
PR	18-JUN-1999;	99US-0139458.	
PR	18-JUN-1999;	99US-0139459.	
PR	18-JUN-1999;	99US-0139460.	

PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.

PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 16.8%; Score 451; DB 21; Length 505;
Best Local Similarity 27.2%; Pred. No. 6.8e-38;
Matches 135; Conservative 104; Mismatches 217; Indels 40; Gaps 16;

QY 39 KKQFGTITIVSLAFVICNSWAGISGSLQLALLAGGPVTLTYGILISTLVYICIAFSLAEL 98
|: : : : | : | : : | | | : | : : | : | :
Db 18 krdlsvfnsfaisfsilsvtgitttyntglrfggtvtlvygwflagstmcvglmsei 77
QY 99 TSVYPTAGQYHFASILAPKINSRISYV--CGLVSLLSWIAIGSSVTMIPAQQIPALI- 155
| | | : | : : : | : | : | | | : | : | :
Db 78 cssyptsgglywsamlagprwvlaswmtgckvvclqw-avtasvdfslaqlivl 136
QY 156 ----AAYSHYTSQDSWHVFLIYEGVALVVLLENLPALKRNPWVHEIG----FGLTIALF 206
| : : | : : | : | : | : : | : :
Db 137 lstggrngggYkgsdfvigihggilfihallnslpisvlsfigqlaalwnllgvlvimi 196
QY 207 VISFTAILARSNPKAPNSQVWTAWSNYTG---WSDGVCFILGLSTSCFMFIGLDAAMHLA 263
: | : | : | : | : | : | : | : | : | : | : | : |

Db 197 liplv-----sterattkfvfntfntdnglqitsyayifvlglmsqytitgydasahmt 251

Qy 264 EECTDAARTVPKAVVSATIIIGFCTAFPPYTIADVIGITDLDLSILS----SAGYIPFFETMQ 319

Db 252 eetvdadkngprgiisaigisilfgwgyilgisavtdipllsetnnsggyalaeifyl 311

Qy 320 SLRSL--SFATVLSGGI--VMAFFALNAVQETASRLTWSFARDNGLVFSSTHLERIHPRW 375

Db 312 afknrfgsqgtgiclvavavfvcgmssvtsnsrmayafsrdegampmsplwhkvnsr- 370

Qy 376 QVPVWSLFATW--GILATCGCI-FLGSSSTAFNALVNSAVVLOQLSFLIPIALLLYQKRDP 432

Db 371 evpi---navwlsalisfcmaltslgsivafqamvsiatiqlyiayaipilrvtiarn- 426

Qy 433 KFLPSTRAFVLPRGIGFLVNVLAVVETSVTTVFESFPLTVPTAASTMNYTSAIIGVALAL 492

Db 427 tfvpg--pfslgk-ygmvgvavlwvvtlsvlfslpvaypitaetlnytpvavaglvai 483

Qy 493 GVLNVTVVHARKHYQGP 508

Db 484 tisywlfsarhwtgtp 499

RESULT 3

AAG31986

ID AAG31986 standard; Protein; 517 AA.

XX

AC AAG31986;

XX

DT 17-OCT-2000 (first entry)

XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 38506.

XX

KW Protein identification; signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

KW termination sequence.

XX

OS Arabidopsis thaliana.

XX

PN EP1033405-A2.

XX

PD 06-SEP-2000.

XX

PF 25-FEB-2000; 2000EP-0301439.

XX

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0130891.

PR 30-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 06-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139492.

PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999; 99US-0139460.

PR 18-JUN-1999; 99US-0139461.

PR 18-JUN-1999; 99US-0139462.

PR 18-JUN-1999; 99US-0139463.

PR 18-JUN-1999; 99US-0139750.

PR 18-JUN-1999; 99US-0139763.

PR 21-JUN-1999; 99US-0139817.

PR 22-JUN-1999; 99US-0139899.

PR 23-JUN-1999; 99US-0140353.

PR 23-JUN-1999; 99US-0140354.

PR 24-JUN-1999; 99US-0140695.

PR 28-JUN-1999; 99US-0140823.

PR 29-JUN-1999; 99US-0140991.

PR 30-JUN-1999; 99US-0141287.

PR 01-JUL-1999; 99US-0141842.

PR 01-JUL-1999; 99US-0142154.

PR 02-JUL-1999; 99US-0142055.

PR 06-JUL-1999; 99US-0142390.

PR 08-JUL-1999; 99US-0142803.

PR 09-JUL-1999; 99US-0142920.

PR 12-JUL-1999; 99US-0142977.

PR 13-JUL-1999; 99US-0143542.

PR 14-JUL-1999; 99US-0143624.

PR 15-JUL-1999; 99US-0144005.

PR 16-JUL-1999; 99US-0144085.

PR 16-JUL-1999; 99US-0144086.

PR 19-JUL-1999; 99US-0144325.

PR 19-JUL-1999; 99US-0144331.

PR 19-JUL-1999; 99US-0144332.

PR 19-JUL-1999; 99US-0144333.

PR 19-JUL-1999; 99US-0144334.

PR 19-JUL-1999; 99US-0144335.

PR 20-JUL-1999; 99US-0144352.

PR 20-JUL-1999; 99US-0144632.

PR 20-JUL-1999; 99US-0144884.

PR 21-JUL-1999; 99US-0144814.

PR 21-JUL-1999; 99US-0145086.

PR 21-JUL-1999; 99US-0145088.

PR 22-JUL-1999; 99US-0145085.

PR 22-JUL-1999; 99US-0145087.

PR 22-JUL-1999; 99US-0145089.

PR 22-JUL-1999; 99US-0145192.

PR 23-JUL-1999; 99US-0145145.

PR 23-JUL-1999; 99US-0145218.

PR 23-JUL-1999; 99US-0145224.

PR 26-JUL-1999; 99US-0145276.

PR 27-JUL-1999; 99US-0145913.

PR 27-JUL-1999; 99US-0145918.

PR 27-JUL-1999; 99US-0145919.

PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.

PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.

PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.
Query Match 16.7%; Score 449.5; DB 21; Length 504;		
Best Local Similarity 26.7%; Pred. No. 9.7e-38;		
Matches 132; Conservative 106; Mismatches 219; Indels 37; Gaps 14;		
Qy	39 KKQFGTITIVSLAFVIONSWAGISGSLQALLAGGPVTLTYGILISTLVYICIAFSLAE 98	
Db	18 krdlsvfnsfaisfsiisvltgttntynglrfggtvtlvvgwflagsftmcvglsmael 77	
Qy	99 TSVYPTAGQYHEASILAPKSIINRSISYVCGLYSLLSWIAIGSSVTMIPAQQIPALI--- 155	
Db	78 cssyptsgglyywsamiagprwaplasmgtwfnivgqavtasvdfslaqlivivlis 137	
Qy	156 --AAYSHTYSQDSWHVFLIYEGVALVLLFNLFALKRNPWWHEIG-----FGLTIALFEVI 208	
Db	138 tggrrngggykgsdfvvgighggilfihallnslpsvlsfiglaalnllglvvlmili 197	
Qy	209 SFTAILARSNPKAPNSQVWTAWSNYTG---WSDGVCFILGLSTSCFMFIGLDAAMHLAE 265	
Db	198 plv-----sterattkfvtnfntdnglgtisyayifvlgllmsgytitgydasahmtee 252	
Qy	266 CTDAARTVPKAVVSAILIGFCTAFPTYTIAVLYGITDLSILS---SAGYIPFETMTQSL 321	
Db	253 tvdadkngprgilsaigisllfgwgylgisyavtdipslsetnnsqgyaiaaeifylaf 312	
Qy	322 RSL--SFATVLSGGI--YMAFFALNAVQETASRLTWSFARDNGLVFSSTHLERIHPRQV 377	
Db	313 knrfsgtggtgiclvvavavvfcgmssvtsnrmayafsrddgampnsplwhkvnsr-ev 371	
Qy	378 PWWSLEATW--GILATCGCI-FLGSSSTAFNALVNSAVWLQQLSFLIPIALLLYQKRDPKF 434	
Db	372 pi---navwlsalisfcmaltslgsivafqamvsiatiglyiaayapilrvltarn-tf 427	
Qy	435 LPSTRAFVLPRIIGFLVNVLAWVFTSVTTVFSPPLTVPSTAASMTMNTVSAIIGVALGV 494	
Db	428 vpg--pfslgk-ygmvvvgwvavlvvvtisvlfpvaypitaetlnytpvavaglvaiti 484	
Qy	495 LNwVVHARKHYQGP 508	
Db	485 sywlfsarhwftgp 498	

RESULT	5
AAG49885	
ID	AAG49885 standard; Protein; 504 AA.
XX	
AC	AAG49885;
XX	
DT	18-OCT-2000 (first entry)
XX	
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 63155.
XX	
KW	Protein identification; signal transduction pathway; metabolic pathway;
KW	hybridisation assay; genetic mapping; gene expression control; promoter;
KW	termination sequence.
XX	
OS	Arabidopsis thaliana.
XX	
PN	EP1033405-A2.
XX	
PD	06-SEP-2000.
XX	
PF	25-FEB-2000; 2000EP-0301439.
XX	
PR	25-FEB-1999; 99US-0121825.
PR	05-MAR-1999; 99US-0123180.
PR	09-MAR-1999; 99US-0123548.
PR	23-MAR-1999; 99US-0125788.
PR	25-MAR-1999; 99US-0126264.
PR	29-MAR-1999; 99US-0126785.
PR	01-APR-1999; 99US-0127462.
PR	06-APR-1999; 99US-0128234.
PR	08-APR-1999; 99US-0128714.
PR	16-APR-1999; 99US-0129845.
PR	19-APR-1999; 99US-0130077.
PR	21-APR-1999; 99US-0130449.
PR	23-APR-1999; 99US-0130510.
PR	23-APR-1999; 99US-0130891.
PR	28-APR-1999; 99US-0131449.
PR	30-APR-1999; 99US-0132048.
PR	30-APR-1999; 99US-0132407.
PR	04-MAY-1999; 99US-0132484.
PR	05-MAY-1999; 99US-0132485.
PR	06-MAY-1999; 99US-0132486.
PR	06-MAY-1999; 99US-0132487.
PR	07-MAY-1999; 99US-0132863.
PR	11-MAY-1999; 99US-0134256.
PR	14-MAY-1999; 99US-0134218.
PR	14-MAY-1999; 99US-0134219.
PR	14-MAY-1999; 99US-0134221.
PR	14-MAY-1999; 99US-0134370.
PR	18-MAY-1999; 99US-0134768.
PR	19-MAY-1999; 99US-0134941.
PR	20-MAY-1999; 99US-0135124.
PR	21-MAY-1999; 99US-0135353.
PR	24-MAY-1999; 99US-0135629.
PR	25-MAY-1999; 99US-0136021.
PR	27-MAY-1999; 99US-0136392.
PR	28-MAY-1999; 99US-0136782.
PR	01-JUN-1999; 99US-0137222.
PR	03-JUN-1999; 99US-0137528.
PR	04-JUN-1999; 99US-0137502.
PR	07-JUN-1999; 99US-0137724.
PR	08-JUN-1999; 99US-0138094.
PR	10-JUN-1999; 99US-0138540.
PR	10-JUN-1999; 99US-0138847.
PR	14-JUN-1999; 99US-0139119.
PR	16-JUN-1999; 99US-0139452.
PR	16-JUN-1999; 99US-0139453.
PR	17-JUN-1999; 99US-0139492.
PR	18-JUN-1999; 99US-0139454.
PR	18-JUN-1999; 99US-0139455.
PR	18-JUN-1999; 99US-0139456.
PR	18-JUN-1999; 99US-0139457.
PR	18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.

PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0153486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 16.7%; Score 449.5; DB 21; Length 504;
Best Local Similarity 26.7%; Pred. No. 9.7e-38;
Matches 132; Conservative 106; Mismatches 219; Indels 37; Gaps 14;

QY 39 KKQFGTITIVSLAFVICSNWAGISGSLQALLAGGPVTLLYGILISTLVYICIAFSLAEL 98
| : : : : | : | : | | (((: : : : | : : | : | :
| : : : : | : | : | | (((: : : : | : : | : | :
Db 18 krdlsvfnsfaisfaislvtgitttntgtlrfggtvtlvvgwflagsftmcvglsmaei 77
| | | | | : : : : | : : | : | : | : | : | : | :
QY 99 TSVYPTAGQYHFASILAPKSINRSISYVCGLVSLLSWIAIGSSVTMIPACQIPAL1--- 155
| | | | | : : : : | : : | : | : | : | : | : | :
Db 78 cssyptsgglyywsamlagprwaplaswmtgwnivgqwtasvdfslaqliqvlls 137
| | | | | : : : : | : : | : | : | : | : | : | :
QY 156 --AAYSHTYSDSWHVFLIEGVALVLLFNLFALKRNPWVHEIG-----FGLTIALFVI 208
| | | | | : : : : | : : | : | : | : | : | : | :
Db 138 tggrrngggykgsdfvvihgghlfihallnslpislvsfigqlaalwnllgvlmll 197
| | | | | : : : : | : : | : | : | : | : | : | :

PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.

PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 16.7%; Score 449.5; DB 21; Length 516;
Best Local Similarity 26.7%; Pred. No. 1e-37;
Matches 132; Conservative 106; Mismatches 219; Indels 37; Gaps 14;

QY 39 KKQFGTITIVSLAFVICNSWAGISGSLQALLAGGPVTLTYGILISTLVYICIAFSLAEL 98
| : : : : : | : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 30 krdlsvsnfaisfslsvltgtitttyntglrfggtvtlvvgwflagstmcvglsmaei 89
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
QY 99 TSVYPTAGGQYHFASILAPKINSRISIVCCGLVSLLSWIAICSSVTMIPAOQIPALI--- 155
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 90 cssyptsgglyywsamlagprwplasmwtgwnivgqwtasvdfslaqliqvivils 149
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
QY 156 --AAYSHTYSDSWHVFLIYEGVALVLLFNLFALKRNPVWHEIG----FGLTIALFVI 208
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 150 tggrrnggykgsdfvvigihggilfihallnslpislvsfigqlaalwnllgvlvjmill 209
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
QY 209 SFIAILARSNPKAPNSQVWTAWSNYTG---WSDGVCFILGLSTSCPMFIGLDAAMHLAEE 265
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 210 plv-----sterattkfvtnfntdnglgtitsyayifvlgllmsgytltgydasahmtee 264
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
QY 266 CTDAARTVPKAVVSAILIIGFCTAFPPYTIAYLYGITOLDLSILS---SAGYIPFETMTQSL 321
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 265 tvdadkngrgilsaiglsilfgwyilglsyavtdipslisetnsggyalaeifylaf 324
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
QY 322 RSL--SFATVLSGGI--VMAFFALNAVQETASRLTWSFARDNGLVFSTHLERIHPRWQV 377
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 325 knrfsggtgclgvavavavfcgmsstsvtsnrmayafardgampmsplwhkvnsr-ev 383
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
QY 378 PVWSLFATW--GILATCGCI--FLGSSPAPNALVNSAVVLOQLSLPIPIALLLYQKRDPKF 434
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 384 pi--navwlsalisfcmaltslgsivafqamvsfatiqlyiaipilrlvtlarn-tf 439
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
QY 435 LPSTRAFLVPRGIGFLVNVLAVVFTSVTTFVFFSPLTVPTAASTMNYTSAIIGVALALGV 494
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 440 vpg--pfsigk-ygmvgvgwvavlvwvvtislsvlslpvaaypltaetlnytpvavaglvaitl 496
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
QY 495 LNWVVHARKHYQGP 508
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 497 sywlfsarhwtgtp 510
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :

RESULT 7
AAG49884
ID AAG49884 standard; Protein; 516 AA.
XX
AC AAG49884;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 63154.
XX
KW protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123160.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132046.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.

PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.

```
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 16.7%; Score 449.5; DB 21; Length 516;
Best Local Similarity 26.7%; Pred. No. 1e-37;
Matches 132; Conservative 106; Mismatches 219; Indels 37; Gaps 14;

QY 39 KKQFGTITIVSLAFVICNSWAGISGSLQALLACGPVTLTYGILISTLVYICIAFSLAEL 98
Db 30 krdlsvfnsfaisfslisvltgitttyntglrfggvtvlvygwfslagftmcvglsmaei 89

QY 99 TSVYPTAGQYHFASILAPKSNRSISYVCGLYSLLSWIAIGSSVTMPAQQIPALI--- 155
Db 90 cssyptsgglyywsamlegprwaplaswmtgwnivgqavtasvdfslagliqvivils 149

QY 156 --AAYSHRYSDSHVFLIYEGVALVLLFNLFALKRNPWVHEIG-----FGLTIALFVI 208
Db 150 tggrrnggykgsdfvlgilghgfihallnslpislvsfigqlaalwnllgvvlmili 209

QY 209 SFIAILARSNPKAPNSQVWTAWSNYTG--WSDGVCFILGLSTSCFMFIGLDAAMHLAEE 265
Db 210 plv-----sterattkfvtnfntdnglgtisyayifvlgllmsqytitgydasahmtee 264

QY 266 CTDAARTVPKAVSAIIIGFCTAFPTYTIAVLGTDLDSILS----SAGYIPFETMTOSL 321
Db 265 tvdadkngprglisalgisilfgwyilgisvavtdipsilsetnsggyaiaefylaf 324

QY 322 RSL--SPATVLSGGI--VMAFFALNAVQETASRLTWSFARDNGLVFSTHLERIHPRWQV 377
Db 325 knrfsgstggivclgvavavfvcgmssvtsnrmayafsrddgampsplwhkvnsr-ev 383

QY 378 PVWSLFATW--GILATCGCI--FLGSSSTAFNALVNSAVVQLQSLFLPIALLLYQKRDPKF 434
Db 384 pi--navwlsaliscmaltslgsivafqamvsiatiglylayalpilrvtlarn-tf 439

QY 435 LPSTRAFLPRGIGFLVNVLAVVVETSVTTVFESFPLTVPTAASTMYTSAIIGVALALGV 494
Db 440 vpg--pfslgk-ygmvvvgvavlvvvvtislvsilpavypitaetinytpvavglvailt 496

QY 495 LNWVWHARKHYQGP 508
Db 497 sywlfsarhwtgtp 510
```

```
RESULT 8
AAG20280
ID AAG20280 standard; Protein; 528 AA.
XX
AC AAG20280;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 22405.
XX
KW protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
```


Db	162	tggrngggykgsdfvvlghgglfihallnslpislvsfigqlaalwnllgvlvmlil	221	PR	14-MAY-1999;	99US-0134219.
QY	209	SFIAILARSNPKAPNSQVWTAWSNYTG---WSDGVCFILGLSTSCFMFIGLDAAMHLAE	265	PR	14-MAY-1999;	99US-0134221.
Db	222	plv-----sterattkfvfnfntdnglgltsyayifvlgllmsqytitgydasahmtee	276	PR	18-MAY-1999;	99US-0134370.
QY	266	CTDAARTVPKAVSAIIGFCTAFPTYTIAVLYGITDLSILS----SAGYIPFETMTQSL	321	PR	18-MAY-1999;	99US-0134768.
Db	277	tvdadkngprgilsaigisilfgwgyligisyavtdipsllsetnnsggyaiaefylaf	336	PR	19-MAY-1999;	99US-0134941.
QY	322	RSI--SFATVLSGGI--VMAFFALNAVQETASRLTWSFARDNGLVFSTHLRIHPRWQV	377	PR	20-MAY-1999;	99US-0135124.
Db	337	knrfsgtgglvclgvavavfvcgmssvtsnsrmayafsrddgampmsplwhkvnsr-ev	395	PR	21-MAY-1999;	99US-0135353.
QY	378	PWWSLFATW--GILATCGCI-FLGSSTAFNALVNSAVVLQOLSFLIPIALLLYOKRDPKF	434	PR	24-MAY-1999;	99US-0135629.
Db	396	pi---navwlsaislscmaltslgsivafqamvsiatiglyiaipilrvtlarn-tf	451	PR	25-MAY-1999;	99US-0136021.
QY	435	LPSTRAFLVPRGIGFLVNVLAUVFTSVTVFFSFPLTVPTAASIMNYTSAIIGVALALGV	494	PR	27-MAY-1999;	99US-0136392.
Db	452	vpg--pfslgk-ygmvgvgvavlvwvtlsvlfslpvaypitaetinytpvavaglvaitl	508	PR	28-MAY-1999;	99US-0136782.
QY	495	LNWVVHARKHYQGP 508		PR	01-JUN-1999;	99US-0137222.
Db	509	sywlfsarhwftgp 522		PR	03-JUN-1999;	99US-0137528.
RESULT 9						
AAG31988						
XX	AAG31988 standard; Protein; 438 AA.					
AC	AAG31988;					
XX						
DT	17-OCT-2000 (first entry)					
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 38508.					
XX						
KW	Protein identification; signal transduction pathway; metabolic pathway;					
KW	hybridisation assay; genetic mapping; gene expression control; promoter;					
XX	termination sequence.					
OS	Arabidopsis thaliana.					
XX						
PN	EP1033405-A2.					
XX						
PD	06-SEP-2000.					
XX						
PF	25-FEB-2000; 2000EP-0301439.					
XX						
PR	25-FEB-1999;	99US-0121825.		PR	13-JUL-1999;	99US-0143542.
PR	05-MAR-1999;	99US-0123180.		PR	14-JUL-1999;	99US-0143624.
PR	09-MAR-1999;	99US-0123548.		PR	15-JUL-1999;	99US-0144005.
PR	23-MAR-1999;	99US-0125788.		PR	16-JUL-1999;	99US-0144085.
PR	25-MAR-1999;	99US-0126264.		PR	16-JUL-1999;	99US-0144086.
PR	29-MAR-1999;	99US-0126785.		PR	19-JUL-1999;	99US-0144325.
PR	01-APR-1999;	99US-0127462.		PR	19-JUL-1999;	99US-0144331.
PR	06-APR-1999;	99US-0128234.		PR	19-JUL-1999;	99US-0144332.
PR	08-APR-1999;	99US-0128714.		PR	19-JUL-1999;	99US-0144333.
PR	16-APR-1999;	99US-0129845.		PR	19-JUL-1999;	99US-0144334.
PR	19-APR-1999;	99US-0130077.		PR	19-JUL-1999;	99US-0144335.
PR	21-APR-1999;	99US-0130449.		PR	20-JUL-1999;	99US-0144352.
PR	23-APR-1999;	99US-0130510.		PR	20-JUL-1999;	99US-0144632.
PR	28-APR-1999;	99US-0130891.		PR	20-JUL-1999;	99US-0144884.
PR	30-APR-1999;	99US-0131449.		PR	21-JUL-1999;	99US-0144814.
PR	30-APR-1999;	99US-0132048.		PR	21-JUL-1999;	99US-0145086.
PR	30-APR-1999;	99US-0132407.		PR	21-JUL-1999;	99US-0145088.
PR	04-MAY-1999;	99US-0132484.		PR	22-JUL-1999;	99US-0145085.
PR	05-MAY-1999;	99US-0132485.		PR	22-JUL-1999;	99US-0145087.
PR	06-MAY-1999;	99US-0132486.		PR	22-JUL-1999;	99US-0145089.
PR	06-MAY-1999;	99US-0132487.		PR	22-JUL-1999;	99US-0145192.
PR	07-MAY-1999;	99US-0132863.		PR	23-JUL-1999;	99US-0145145.
PR	11-MAY-1999;	99US-0134256.		PR	23-JUL-1999;	99US-0145218.
PR	14-MAY-1999;	99US-0134218.		PR	23-JUL-1999;	99US-0145224.

PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.

PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.

PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.
Query Match 14.8%; Score 398.5; DB 21; Length 437;		
Best Local Similarity 26.8%; Pred. No. 1.6e-32;		
Matches 119; Conservative 95; Mismatches 193; Indels 37; Gaps 14;		
Qy	89	ICIAFSLAELTSVYPTAGQYHFASILAPKSIKRSISYVCGLVSLLSWIAIGSSVTMIPA 148
Db	1	mcvglsmaeicssyptsgglyywsamlagprwaplasmwtgwnivgqwavtasvdfsia 60
Qy	149	QQIPALI----AAVSHYTSQDSWHVFLIYEGVALVLLFNLFALKRNPWVHEIG---- 198
Db	61	qliqvllstggrngggykgsdfvviighgilfihallnslpislvsfigqlaalwnl 120
Qy	199	FGLTIALFVISFIALAKSNPKAPNSQVWTAWSNYTG--WSDGVCFILGLSTSCPFMFIG 255
Db	121	lgvilmiliplv----sterattkfvftntdnglitsyayifvlgllmsqytilg 175
Qy	256	LDAAMHLAEECTDAARTVPKAVVSAIIIGFCTAFPPYTIADVLYGITDLDSILS----SAGY 311
Db	176	ydasahmteetvdadkngprgiisaigisilfgwgylgisayavtdipslsetnnsgy 235
Qy	312	IPFETMTQSLRSL--SFATVLSGGGI--VMAPPALNAVQETASRLTWSFARDNGLVFSTH 367
Db	236	aiaaeifylafknrfsgtggtgiclvavavfcgmssvtsnsrmayafsrddgampmspl 295
Qy	368	LERIHPRWQVPVWSLEATW--GILATCGCI-FLGSSSTAFNALVNSAVVLQQLSFLIPIAL 424
Db	296	whkvnsl-evpi--navwlsalisfcmaltsgisvafqamvsiatiglyiayaiipil 351
Qy	425	LLYQKRPKFLPLPSTRAFLPRGIGFLVNVLAVVTSTVTVFFSFPLTVPTAASTMNYTSA 484
Db	352	rvtlarn-tfvpvg--pfslgk-ygmvgvgwvavlvwvtislvsflslpvaypitaelnlytpv 407
Qy	485	IIGVALALGVLNWWVHARKHYQGP 508
Db	408	avaglvaitlsywlfsarhwftgp 431
RESULT 11		
AAG49911		
ID	AAG49911 standard; Protein; 439 AA.	
XX		

AC	AAG49911;	
XX	18-OCT-2000	(first entry)
DT	Arabidopsis thaliana	protein fragment SEQ ID NO: 63191.
DE	Protein identification; signal transduction pathway; metabolic pathway;	
XX	hybridisation assay; genetic mapping; gene expression control; promoter;	
KW	termination sequence.	
KW	Arabidopsis thaliana.	
XX	EP1033405-A2.	
OS	06-SEP-2000.	
PN	25-FEB-2000;	2000EP-0301439.
XX	25-FEB-1999;	99US-0121825.
PR	05-MAR-1999;	99US-0123180.
PR	09-MAR-1999;	99US-0123548.
PR	23-MAR-1999;	99US-0125788.
PR	25-MAR-1999;	99US-0126264.
PR	29-MAR-1999;	99US-0126785.
PR	01-APR-1999;	99US-0127462.
PR	06-APR-1999;	99US-0128234.
PR	08-APR-1999;	99US-0128714.
PR	16-APR-1999;	99US-0129845.
PR	19-APR-1999;	99US-0130077.
PR	21-APR-1999;	99US-0130449.
PR	23-APR-1999;	99US-0130510.
PR	23-APR-1999;	99US-0130891.
PR	28-APR-1999;	99US-0131449.
PR	30-APR-1999;	99US-0132048.
PR	30-APR-1999;	99US-0132407.
PR	04-MAY-1999;	99US-0132484.
PR	05-MAY-1999;	99US-0132485.
PR	06-MAY-1999;	99US-0132486.
PR	06-MAY-1999;	99US-0132487.
PR	07-MAY-1999;	99US-0132863.
PR	11-MAY-1999;	99US-0134256.
PR	14-MAY-1999;	99US-0134218.
PR	14-MAY-1999;	99US-0134219.
PR	14-MAY-1999;	99US-0134221.
PR	14-MAY-1999;	99US-0134370.
PR	18-MAY-1999;	99US-0134768.
PR	19-MAY-1999;	99US-0134941.
PR	20-MAY-1999;	99US-0135124.
PR	21-MAY-1999;	99US-0135353.
PR	24-MAY-1999;	99US-0135629.
PR	25-MAY-1999;	99US-0136021.
PR	27-MAY-1999;	99US-0136392.
PR	28-MAY-1999;	99US-0136782.
PR	01-JUN-1999;	99US-0137222.
PR	03-JUN-1999;	99US-0137528.
PR	04-JUN-1999;	99US-0137502.
PR	07-JUN-1999;	99US-0137724.
PR	08-JUN-1999;	99US-0138094.
PR	10-JUN-1999;	99US-0138540.
PR	10-JUN-1999;	99US-0138847.
PR	14-JUN-1999;	99US-0139119.
PR	16-JUN-1999;	99US-0139452.
PR	16-JUN-1999;	99US-0139453.
PR	17-JUN-1999;	99US-0139492.
PR	18-JUN-1999;	99US-0139454.
PR	18-JUN-1999;	99US-0139455.
PR	18-JUN-1999;	99US-0139456.
PR	18-JUN-1999;	99US-0139457.
PR	18-JUN-1999;	99US-0139458.
PR	18-JUN-1999;	99US-0139459.
PR	18-JUN-1999;	99US-0139460.
PR	18-JUN-1999;	99US-0139461.
PR	18-JUN-1999;	99US-0139462.

PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 12.1%; Score 326.5; DB 21; Length 521;
Best Local Similarity 26.0%; Pred. NO. 6.4e-25;
Matches 114; Conservative 83; Mismatches 172; Indels 69; Gaps 17;
QY 130 LVSLLSWIAIGSSVTMIPAQQIPALIAAYSHYTSQDS-----WHV-----FLIYEGV 176
|: : |: : |: : |: : |: : |: : |: : |: : |: : |: : |: : |: : |: : |: : |: :
Db 88 llhnrvwlygrdlllpylrwsllecyarwpslgsscllddlwlvavtasvdfslaqli 147
QY 177 ALVVLL-----FNLFALKRN-PWVHEIGFELTIALFVISFIAILAR----- 216
:
Db 148 qvillstggrnggykgsdfvvihihgglfihallnsipis--vlsfigqlaalwnll 205
QY 217 -----SNPKAPNSQVWTAWSNYTG---WSDGVCFILGLSTSCFMFIGLDAAMH 261
|: : |: : |: : |: : |: : |: : |: : |: : |: : |: : |: : |: : |: : |: : |: :
Db 206 gvlvmlilplvsterattkfvtnfntdnglitsyavifvlgllmsqytitgydasah 265
QY 262 LAECTDAARTVPKAVVSAIIIGFCTAFPTYTIAVLYGITDLSILS-----SAGYIPFETM 317
:
Db 266 mteetvdadkngprgilsaigisilfgwyilgisavtdipslisetnnsqgyaiaelf 325
QY 318 TQSLRSL--SFATVLSGGI--VMAFFALNAVQETASRLTWSFARDNGLVSTHLRIHP 373
: : : |: : |: : |: : |: : |: : |: : |: : |: : |: : |: : |: : |: : |: : |: :
Db 326 ylafknrfgsgtgglvclgvavavfvcgmssvtsnrmayafsdgampmsplwhkvns 385
QY 374 RWQVPVWSLFATW--GILATCGCI-FLGSSTAFNALVNSAVVLQQLSFLIPIALLLYQKR 430
|: : |: : |: : |: : |: : |: : |: : |: : |: : |: : |: : |: : |: : |: : |: :
Db 386 r-evpi--navwlsalisfcmaltslgsivafqamvsiatiglyayaipilrvltlar 441
QY 431 DPKFLPSTRAFLVPRGIGFLVNVLAUVFTSVTTVEFFSPFLTVPTAASTMNYTSAIIGVAL 490
: : : |: : |: : |: : |: : |: : |: : |: : |: : |: : |: : |: : |: : |: : |: :
Db 442 n-tfvpq--pfslgk-ygmvvvgwvavlvwvvtisvlfsipvaypitaetlnytpvavaglv 497
QY 491 ALGVLNWWVHARKHYQGP 508
|: : |: : |: : |: : |: : |: : |: : |: : |: : |: : |: : |: : |: : |: : |: :
Db 498 aitlsywlfsarhwftgp 515
RESULT 13
AAG49912
ID AAG49912 standard; Protein; 311 AA.
XX
AC AAG49912;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 63192.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.

PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.

PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 11.3%; Score 305; DB 21; Length 311;
Best Local Similarity 29.4%; Pred. No. 5.2e-23;
Matches 90; Conservative 62; Mismatches 132; Indels 22; Gaps 11;

QY 217 SNPKAPNSQVWTANSNYTG--WSDGVCFILGLSTSCFMFIGLDAAMHLAEECTDAARTV 273
Db 8 sterattkfvfntdnglgltsyavifvlgllmsqytitgydasahmtteetvdadkng 67
QY 274 PRAVVSIIIGFCTAPPYTTIAVLGIDTLDLSILS----SAGYIPFETMTQSLRSL--SFA 327
Db 68 prgiisaigisllifwgvyllgisavtdipslsetnnsqgyaiaelfylafknrfsgt 127
QY 328 TVLSCCGI--VMAPFALNAVQETASRLTWSFARDNGLVFTSLHRIHPRWQVPVWSLEAT 385
Db 128 ggivclgvavavfegmssvtsnrmayafsrddgampmsplwhkvnsr-evpi---nav 183
QY 386 W--GILATCGCI-FLGSSTAFNALVNSAVVLOQLSFLIPTALLLYQKRPKFLPSTRAFV 442
Db 184 wlsalisfcmaltsgsivafqamvsiatiglyiayalpilrvtlarn-tfvpqg--pfs 240
QY 443 LPRGIGFLVNLVAVFTSVTTVFFSFPLTVPTAASTMNVTSAIIGVALGVLNWNVHAR 502
Db 241 lqk-ygmvgvwavllwvvtisvlfsalpvaypitaetlnytpvavaglvaitlsywlfsar 299
QY 503 KHYQGP 508
Db 300 hwftgp 305

RESULT 14
AAU37348
ID AAU37348 standard; Protein: 482 AA.

AC AAU37348;
XX
DT 13-FEB-2002 (first entry)
XX
DE Staphylococcus aureus cellular proliferation protein #1518.
XX
KW Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX
OS Staphylococcus aureus.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.

PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI; 2001-611495/70.
DR N-PSDB; AAS55207.
XX
PT New polynucleotides for the identification and development of
XX antibiotics, comprise sequences of antisense nucleic acids -
PS Example 3; Seq ID No 12941; 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 482 AA;

Query Match 8.0%; Score 216; DB 22; Length 482;
Best Local Similarity 24.4%; Pred. No. 1.8e-13;
Matches 132; Conservative 79; Mismatches 209; Indels 120; Gaps 29;
QY 5 PSYGCEKGGTQRTTKNTETAAGGASESLNPLEKKQFGTITIVSLAFVICSNWAGISGS 64
Db 14 palygnkdghlkrtrvrdflalgv-----gtivstsl-ftl----pgl--- 52
QY 65 LQALLAGGPVTLGYICIAFSLAELTSVYPTAGGQYHFASILAPKSINRSI 124
Db 53 --vaaehagpaval-sflaaivaglvafvtaemaampfagsaywvnl----fgeff 105
QY 125 SYVCGLVSL-----LSWIAIGSSVTM-----IPAQQIPALIAAYSHYTSQDSWH 168
Db 106 gwvagwallaeyfiavafvasgfsanlrglvkplgielp-----aalsnpgftnggf 157
QY 169 VELIYEGVALVLLFNLFALKRNPWVHEIGFC-----LTIALFVISFIAL 214
Db 158 idil---aaivilltallsr-----gmseaarmentlilvklailflvivgltai 206
QY 215 ARSN--PKAPNSQVWTANSNYTGWSDGVCFIGLSTSCFMFIGLDAAMHLAEECTDAART 272
Db 207 nvsnyvpfipekfv-tatgdfggwqg---iyagvsmiflayigfidsiaansaealdpqt 262
QY 273 VPKAVVSAIIIGFCTAPPYTTIAVLGIDTLDLSILSAGYIPFETMTQSLRSLS---FATV 329
Db 263 mprgilgslsvai-vlfiavavlvvgmfhygqyanna-----epvgwalrqsgghgvvaa 316
QY 330 LSCGGIVMAFFALNAVQETASRLTWSFARDNGLVFS--THLERIHPRWQVPVWSLFATWG 387

Db 317 vqaisvlgmftaligmnlagsrlllysfyrd-gllpswlslnkdh----lpral----- 366

Qy 388 ILATCGCIEFGS--SPAFNALVNSAVVLQQLSFLIPIALLLYQKRDPKFLPSTRAFVLPR 445

Db 367 vltiigvligsmfpaflaqlisagtivafmf-vslamyrlrkregkdlp-ipafklp- 423

Qy 446 GIGFLVNL-AVVFSTVTTFVFFSFPLTVPTAASTMNYSALIGVALALGVLMNVVHARKH 504

Db 424 ----lypvlpaitfvllvfwglgf----eaklytlwifvigiilylsy--girhskkn 473

RESULT 15

ID AAG90272

XX AAG90272 standard; Protein; 453 AA.

AC AAG90272;

XX 26-SEP-2001 (first entry)

XX C glutamicum protein fragment SEQ ID NO: 4026.

DE C glutamicum protein fragment SEQ ID NO: 4026.

XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;

KW organic acid synthesis.

KW organic acid synthesis.

XX Corynebacterium glutamicum.

OS Corynebacterium glutamicum.

XX EP1108790-A2.

PN EP1108790-A2.

XX 20-JUN-2001.

PD 20-JUN-2001.

XX 18-DEC-2000; 2000EP-0127688.

PF 18-DEC-2000; 2000EP-0127688.

XX 16-DEC-1999; 99JP-0377484.

PR 07-APR-2000; 2000JP-0159162.

PR 03-AUG-2000; 2000JP-0280988.

XX (KYOW) KYOWA HAKKO KOGYO KK.

PA Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

XX Tateishi N, Senoh A, Ikeda M, Ozaki A;

PI WPI; 2001-376931/40.

DR N-PSDB; AAH65491.

DR Novel polynucleotides derived from Coryneform bacteria, for identifying

PT mutation point of a gene, measuring expression of a gene, analysing

PT expression profile or pattern of a gene and identifying homologous gene

PT expression profile or pattern of a gene and identifying homologous gene

XX Claim 17; SEQ ID NO: 4026; 246pp + Sequence Listing; English.

PS The present invention provides a number of nucleotide and protein

XX sequences from the Coryneform bacterium Corynebacterium glutamicum. These

CC are useful for identifying the mutation point of a gene derived from a

CC mutant of coryneform bacterium, measuring expression amount and

CC analysing the expression profile or expression pattern of a gene derived

CC from Coryneform bacterium, and identifying a homologue of a gene derived

CC from coryneform bacterium. Coryneform bacteria are useful for producing

CC amino acids, nucleic acids, vitamins, saccharides and organic acids,

CC particularly L-lysine. The present sequence is a protein described

CC in the exemplification of the invention.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from the

CC European Patent Office.

XX Sequence 453 AA;

SQ

Query Match 7.6%; Score 204.5; DB 22; Length 453;

Best Local Similarity 23.7%; Pred. No. 2.5e-12;

Matches 111; Conservative 89; Mismatches 181; Indels 87; Gaps 27;

Qy 63 GSLQALLAGGPVTLTYGILISTLVYICIAFSLAELTSVYPTAGGQYHFASILAPKSINR 122

Db 22 gsgktsvagspsvllvyailgfmffvnmram-----gelllan-lykkslrd 67

Qy 123 SIS-----YVCGLVSLLSWIAIGSSVTMIPAAQOIPALIAAYSHYTSQDSW-HVFLIY 173

Db 68 avsdilpggagfvgtwyfcwiatgma-----divaitg----ytqywwpeiplwl 115

Qy 174 EGVALVLLF--NLFAIK---RNPWVHEIGFGLTIALFVIS-FIAILARSNPKAPNSQV 226

Db 116 pgvltiallfalnlaavrlfgemefwfaikivaivslivvglfmvtafespngttaqf 175

Qy 227 WTAWSNYTGWSDGVC-FILGLSTSCFMFICGLDAAMHMLAEECTDAARTVPKAVVSAIIGF 285

Db 176 nnliehggffpogitgflagfqiaifavgielagtaaaetenptktlpra-insipiri 234

Qy 286 CTAFPPYTI AVLYGITDLDSDJLSSAGYIPFETMTQSLRSLSEAFVLSGCGIVMAFFALNAV 345

Db 235 vvfylalavimmvtpwdqv--radnspfvqm-----falagipaaagli-nfvvitsa 285

Qy 346 QETASRLTWSEFARDNGLVFSTHLERHP-RWQ-----VPVWSL-FATWGILATCGCIFL 397

Db 286 assansglfstsr--mlyglslegaaaprwsrlsknlvpargltfsviclipavgliya 342

Qy 398 GSST--AFNAL--VNSAVVLQQLSFLIPIALLLYQKRDPKFLPSTRAFVLPRGIGFLVNV 453

Db 343 ggtvieaftlittvssvlfmwwsyil-vayivyrnspe-lhkksifkmpggv-----v 395

Qy 454 LAVVFTSVTTVFFSFPLTV----PTAASTMNYTSA---IIGVA-LALG 493

Db 396 mav----vvlvffaamlvvlslslepdtraaliatpvpwfillgigwlsig 439

Search completed: April 27, 2002, 07:45:35

Job time: 8116 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2002, 20:50:14 ; Search time 5001.19 Seconds
(without alignments)
4258.625 Million cell updates/sec

Title: US-09-882-694A-7
Perfect score: 1578
Sequence: 1 atggactccagaccaagtgg.....aatttcaagttgggccatga 1578

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	55	3.5	1010	12 CNS06EC5	AL394923 T7 end of
2	54	3.4	521	12 AZ928850	AZ928850 479.dif18
c 3	51.2	3.2	939	12 CNS07DIC	AL440506 T7 end of
4	49.8	3.2	686	10 BI954357	BI954357 HVSMEM001
5	49.6	3.1	634	9 AW442992	AW442992 EST307922
c 6	49	3.1	948	12 CNS06DGC	AL393770 T7 end of
7	48.8	3.1	525	12 AQ397336	AQ397336 mgxb0012D
c 8	46.2	2.9	644	12 AQ161795	AQ161795 mgxb00091
c 9	43	2.7	1028	12 CNS06XLB	AL419877 T3 end of
10	42.6	2.7	511	10 BG241224	BG241224 OV1_39_F1
11	42.6	2.7	546	10 BF586115	BF586115 FM1_26_H0
12	42.6	2.7	997	12 CNS005TE	AL060767 Drosophil
c 13	42.4	2.7	375	9 AA784308	AA784308 d4cl2al.f
14	42.2	2.7	679	10 BG810178	BG810178 mgct003xe
15	41.6	2.6	385	9 AU031682	AU031682 AU031682
16	41.6	2.6	675	10 BI954997	BI954997 HVSMEM002
17	41	2.6	601	10 BG463370	BG463370 EM1_48_F1

18	41	2.6	663	10 BI955572	BI955572 HVSMEM002
c 19	40.6	2.6	641	12 CNS077NX	AL432931 T7 end of
c 20	40.2	2.5	1101	12 CNS012S8	AL101954 Drosophil
c 21	39.8	2.5	1273	12 AQ896271	AQ896271 HS_3176_B
c 22	39.6	2.5	969	12 CNS02S69	AL211482 Tetraodon
23	39.2	2.5	383	10 BE523754	BE523754 M41C3STM
24	39.2	2.5	443	9 AA788013	AA788013 r4a12al.r
25	39.2	2.5	450	9 AV548940	AV548940 AV548940
26	37.4	2.4	399	9 AW872351	AW872351 hm28cl1.x
c 27	37.4	2.4	639	9 AL648272	AL648272 AL648272
c 28	37.2	2.4	298	10 BE936629	BE936629 QV4-OT003
c 29	37	2.3	913	12 CNS079UV	AL435773 T3 end of
30	36.8	2.3	517	12 AZ930479	AZ930479 474.dhz54
31	36.8	2.3	748	10 BI970431	BI970431 GM830010B
c 32	36.6	2.3	256	9 BB449103	BB449103 BB449103
c 33	36.6	2.3	490	10 BI773560	BI773560 rp03b12.y
34	36.6	2.3	586	10 BF255233	BF255233 HVSMef000
35	36.6	2.3	956	12 CNS01KTF	AL148772 Anopheles
c 36	36.4	2.3	548	9 AI660809	AI660809 we69all.x
37	36.4	2.3	548	9 AA477575	AA477575 zu44b02.r
38	36.4	2.3	793	12 BH111627	BH111627 RPCI-24-2
39	36.4	2.3	1650	12 AG150948	AG150948 Pan trogl
40	36.2	2.3	328	10 D23882	D23882 R1CR0503A R
c 41	36.2	2.3	469	9 AA655635	AA655635 vs46h11.r
c 42	36.2	2.3	867	12 AZ126336	AZ126336 OSJNBb007
43	36	2.3	394	10 BI776502	BI776502 EBp105_SQ
44	36	2.3	408	9 AW397789	AW397789 sg68d06.y
45	36	2.3	472	9 AL378484	AL378484 MEBB38E05

ALIGNMENTS

RESULT 1	CNS06EC5/c	CNS06EC5	1010 bp	DNA	linear	GSS 17-JUN-2001
LOCUS	T7 end of clone AR0AA020D08 of library AR0AA	T7 end of clone AR0AA020D08 of library AR0AA	from strain CBS 732 of			
DEFINITION	Zygosaccharomyces rouxii, genomic survey sequence.	Zygosaccharomyces rouxii, genomic survey sequence.				
ACCESSION	AL394923	AL394923.1	GI:12146190			
VERSION						
KEYWORDS	GSS.					
SOURCE	Zygosaccharomyces rouxii.					
ORGANISM	Zygosaccharomyces rouxii					
REFERENCE	1 (bases 1 to 1010)					
AUTHORS	de Montigny,J., Straub,M., Potier,S., Tekala,F., Dujon,B., Wincker,P., Artiguenave,F. and Souciet,J.					
TITLE	Genomic exploration of the hemiascomycetous yeasts: 8.					
JOURNAL	Zygosaccharomyces rouxii					
MEDLINE	FEBS Lett. 487 (1), 52-55 (2000)					
AUTHORS	2 (bases 1 to 1010)					
	Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G., Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S., de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B., Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S., Saurin,W., Tekala,F., Toffano-Nloche,C., Wesolowski-Louvel,M., Wincker,P. and Weissenbach,J.					
TITLE	Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies					
JOURNAL	FEBS Lett. 487 (1), 3-12 (2000)					
MEDLINE	20584711					
REFERENCE	3 (bases 1 to 1010)					
AUTHORS	Genoscope.					
TITLE	Direct Submission					
JOURNAL	Submitted (06-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)					
COMMENT	This GSS is part of a random sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces					

```

; APPLICANT: Paoni, Nicholas F.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2830PLC3
; CURRENT APPLICATION NUMBER: US/10/006,063A
; CURRENT FILING DATE: 2002-03-15
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 477
; SEQ ID NO 51
; LENGTH: 1734
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-006-063A-51

Query Match          2.1%; Score 32.6; DB 6; Length 1734;
Best Local Similarity 55.9%; Pred. No. 3;
Matches 62; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 698 atactggctggtccgacggcgctgtgcttcatctcctggcctttcgacatcctgcttcagt 757
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1229 ATCCAGATTCCCGCTCCCGCGGCTTCATTCCCTGGCTTTTCACACCCGGGTTTATGT 1170

QY 758 tcattggcttgacgcagcaatgcattcgtgctgaagaatgcacagatgctg 808
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1169 CCATTTCCTCCCGCTCCACCCGGTGTGCCGGAGGAGCGGTGCTG 1119

RESULT 3
US-10-006-117A-51/c
; Sequence 51, Application US/10006117A
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan I.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2830PLC13
; CURRENT APPLICATION NUMBER: US/10/006,117A
; CURRENT FILING DATE: 2002-03-19
; Prior Application removed - See File Wrapper or Palm
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 477
; SEQ ID NO 51
; LENGTH: 1734
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-006-117A-51

Query Match          2.1%; Score 32.6; DB 6; Length 1734;
Best Local Similarity 55.9%; Pred. No. 3;
Matches 62; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 698 atactggctggtccgacggcgctgtgcttcatctcctggcctttcgacatcctgcttcagt 757
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1229 ATCCAGATTCCCGCTCCCGCGGCTTCATTCCCTGGCTTTTCACACCCGGGTTTATGT 1170

QY 758 tcattggcttgacgcagcaatgcattcgtgctgaagaatgcacagatgctg 808
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1169 CCATTTCCTCCCGCTCCACCCGGTGTGCCGGAGGAGCGGTGCTG 1119
```

```

RESULT 4
US-10-006-130A-51/c
; Sequence 51, Application US/10006130A
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan I.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2830PIC7
; CURRENT APPLICATION NUMBER: US/10/006,130A
; CURRENT FILING DATE: 2002-03-19
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 477
; SEQ ID NO 51
; LENGTH: 1734
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-006-130A-51

Query Match          2.1%; Score 32.6; DB 6; Length 1734;
Best Local Similarity 55.9%; Pred. No. 3;
Matches 62; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 698 atactggctggtccgacggcgctgtgcttcatctcctggcctttcgacatcctgcttcagt 757
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1229 ATCCAGATTCCCGCTCCCGCGGCTTCATTCCCTGGCTTTTCACACCCGGGTTTATGT 1170

QY 758 tcattggcttgacgcagcaatgcattcgtgctgaagaatgcacagatgctg 808
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1169 CCATTTCCTCCCGCTCCACCCGGTGTGCCGGAGGAGCGGTGCTG 1119

RESULT 5
US-10-006-172A-51/c
; Sequence 51, Application US/10006172A
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan I.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2830PIC11
; CURRENT APPLICATION NUMBER: US/10/006,172A
; CURRENT FILING DATE: 2002-03-19
; PRIOR APPLICATION NUMBER: 60/098716
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098723
; PRIOR FILING DATE: 1998-09-01
```


Best Local Similarity 55.9%; Pred. No. 3;
Matches 62; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

Qy 698 atactggcttggtcgcagcgcgctgtctcttcaccccttgcgacatccctgcttcgt 757
 ||| || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1229 ATCCAGATTCCCGCTCCCGGGGCTTCATCCCTGGCTTTTCACACCCGGTTATGT 1170

Qy 758 tcattggcttgagcagcaatgcattggctgaagaatgcacagatgctg 808
 |||| || ||| || || ||| || ||| ||| || ||| |||
Db 1169 CCATTCTCCGCGCGCTCCACCGTGGTTCGCGGAGGAGGACCGGTGCTG 1119

RESULT 8
US-10-017-610A-51/c
; Sequence 51, Application US/10017610A
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan I.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2830PLC64
; CURRENT APPLICATION NUMBER: US/10/017,610A
; CURRENT FILING DATE: 2001-12-13
; PRIOR APPLICATION NUMBER: 60/098716
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098723
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098749
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098750
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098803
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/098821
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/098843
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/099536
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099596
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099598
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099602
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099642
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099741
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099754
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099763
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099792
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099808
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099812
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099815
; PRIOR FILING DATE: 1998-09-10

; PRIOR APPLICATION NUMBER: 60/099816
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/100385
; PRIOR FILING DATE: 1998-09-15
; PRIOR APPLICATION NUMBER: 60/100388
; PRIOR FILING DATE: 1998-09-15
; PRIOR APPLICATION NUMBER: 60/100390
; PRIOR FILING DATE: 1998-09-15
; PRIOR APPLICATION NUMBER: 60/100584
; PRIOR FILING DATE: 1998-09-16
; PRIOR APPLICATION NUMBER: 60/100627
; PRIOR FILING DATE: 1998-09-16
; PRIOR APPLICATION NUMBER: 60/100661
; PRIOR FILING DATE: 1998-09-16
; PRIOR APPLICATION NUMBER: 60/100662
; PRIOR FILING DATE: 1998-09-16
; PRIOR APPLICATION NUMBER: 60/100664
; PRIOR FILING DATE: 1998-09-16
; PRIOR APPLICATION NUMBER: 60/100683
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: 60/100684
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: 60/100710
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: 60/100711
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: 60/100848
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 60/100849
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 60/100919
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: 60/100930
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: 60/101014
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 60/101068
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 60/101071
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 60/101279
; PRIOR FILING DATE: 1998-09-22
; PRIOR APPLICATION NUMBER: 60/101471
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101472
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101474
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101475
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101476
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101477
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101479
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101738
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/101741
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/101743
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/101915
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/101916
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/102207
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/102240
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/102307
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/102330

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2002, 01:33:51 ; Search time 8172.36 Seconds
(without alignments)
4177.480 Million cell updates/sec

Title: US-09-882-694A-7
Perfect score: 1578
Sequence: 1 atggactccagaccagtg.....aattcaagttggccatga 1578

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 21979536 seqs, 10817449327 residues 43959072

Total number of hits satisfying chosen parameters: 43959072

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_Main.*
1: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq.*
2: /cgn2_6/ptodata/2/pna/US06_COMB.seq.*
3: /cgn2_6/ptodata/2/pna/US07_COMB.seq.*
4: /cgn2_6/ptodata/2/pna/US080_COMB.seq.*
5: /cgn2_6/ptodata/2/pna/US081_COMB.seq.*
6: /cgn2_6/ptodata/2/pna/US082_COMB.seq.*
7: /cgn2_6/ptodata/2/pna/US083_COMB.seq.*
8: /cgn2_6/ptodata/2/pna/US084_COMB.seq.*
9: /cgn2_6/ptodata/2/pna/US085_COMB.seq.*
10: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
11: /cgn2_6/ptodata/2/pna/US087_COMB.seq.*
12: /cgn2_6/ptodata/2/pna/US088_COMB.seq.*
13: /cgn2_6/ptodata/2/pna/US089_COMB.seq.*
14: /cgn2_6/ptodata/2/pna/US090_COMB.seq.*
15: /cgn2_6/ptodata/2/pna/US091_COMB.seq.*
16: /cgn2_6/ptodata/2/pna/US092_COMB.seq.*
17: /cgn2_6/ptodata/2/pna/US093_COMB.seq.*
18: /cgn2_6/ptodata/2/pna/US094_COMB.seq.*
19: /cgn2_6/ptodata/2/pna/US095A_COMB.seq.*
20: /cgn2_6/ptodata/2/pna/US095B_COMB.seq.*
21: /cgn2_6/ptodata/2/pna/US095C_COMB.seq.*
22: /cgn2_6/ptodata/2/pna/US095D_COMB.seq.*
23: /cgn2_6/ptodata/2/pna/US096A_COMB.seq.*
24: /cgn2_6/ptodata/2/pna/US096B_COMB.seq.*
25: /cgn2_6/ptodata/2/pna/US096C_COMB.seq.*
26: /cgn2_6/ptodata/2/pna/US096D_COMB.seq.*
27: /cgn2_6/ptodata/2/pna/US096E_COMB.seq.*
28: /cgn2_6/ptodata/2/pna/US097A_COMB.seq.*
29: /cgn2_6/ptodata/2/pna/US097B_COMB.seq.*
30: /cgn2_6/ptodata/2/pna/US097C_COMB.seq.*
31: /cgn2_6/ptodata/2/pna/US098A_COMB.seq.*
32: /cgn2_6/ptodata/2/pna/US098B_COMB.seq.*
33: /cgn2_6/ptodata/2/pna/US098C_COMB.seq.*
34: /cgn2_6/ptodata/2/pna/US099A_COMB.seq.*
35: /cgn2_6/ptodata/2/pna/US099B_COMB.seq.*
36: /cgn2_6/ptodata/2/pna/US099C_COMB.seq.*
37: /cgn2_6/ptodata/2/pna/US100_COMB.seq.*
38: /cgn2_6/ptodata/2/pna/US101_COMB.seq.*
39: /cgn2_6/ptodata/2/pna/US6000_COMB.seq.*
40: /cgn2_6/ptodata/2/pna/US6001_COMB.seq.*
41: /cgn2_6/ptodata/2/pna/US6002_COMB.seq.*
42: /cgn2_6/ptodata/2/pna/US6003_COMB.seq.*
43: /cgn2_6/ptodata/2/pna/US6004_COMB.seq.*

44: /cgn2_6/ptodata/2/pna/US6005_COMB.seq.*
45: /cgn2_6/ptodata/2/pna/US6006_COMB.seq.*
46: /cgn2_6/ptodata/2/pna/US6007_COMB.seq.*
47: /cgn2_6/ptodata/2/pna/US6008_COMB.seq.*
48: /cgn2_6/ptodata/2/pna/US6009_COMB.seq.*
49: /cgn2_6/ptodata/2/pna/US6010_COMB.seq.*
50: /cgn2_6/ptodata/2/pna/US6011_COMB.seq.*
51: /cgn2_6/ptodata/2/pna/US6012_COMB.seq.*
52: /cgn2_6/ptodata/2/pna/US6013_COMB.seq.*
53: /cgn2_6/ptodata/2/pna/US6014_COMB.seq.*
54: /cgn2_6/ptodata/2/pna/US6015_COMB.seq.*
55: /cgn2_6/ptodata/2/pna/US6016_COMB.seq.*
56: /cgn2_6/ptodata/2/pna/US6017_COMB.seq.*
57: /cgn2_6/ptodata/2/pna/US6018_COMB.seq.*
58: /cgn2_6/ptodata/2/pna/US6019_COMB.seq.*
59: /cgn2_6/ptodata/2/pna/US6020_COMB.seq.*
60: /cgn2_6/ptodata/2/pna/US6021_COMB.seq.*
61: /cgn2_6/ptodata/2/pna/US6022_COMB.seq.*
62: /cgn2_6/ptodata/2/pna/US6023_COMB.seq.*
63: /cgn2_6/ptodata/2/pna/US6024_COMB.seq.*
64: /cgn2_6/ptodata/2/pna/US6025_COMB.seq.*
65: /cgn2_6/ptodata/2/pna/US6026_COMB.seq.*
66: /cgn2_6/ptodata/2/pna/US6027_COMB.seq.*
67: /cgn2_6/ptodata/2/pna/US6028_COMB.seq.*
68: /cgn2_6/ptodata/2/pna/US6029_COMB.seq.*
69: /cgn2_6/ptodata/2/pna/US6030_COMB.seq.*
70: /cgn2_6/ptodata/2/pna/US6031_COMB.seq.*
71: /cgn2_6/ptodata/2/pna/US6032_COMB.seq.*
72: /cgn2_6/ptodata/2/pna/US6033_COMB.seq.*
73: /cgn2_6/ptodata/2/pna/US6034_COMB.seq.*
74: /cgn2_6/ptodata/2/pna/US6035_COMB.seq.*
75: /cgn2_6/ptodata/2/pna/US6036_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1578	100.0	1578	17	US-09-351-224E-7	Sequence 7, Appli
2	1578	100.0	1578	26	US-09-677-488A-7	Sequence 7, Appli
3	1578	100.0	1578	26	US-09-677-682A-7	Sequence 7, Appli
4	1578	100.0	1578	26	US-09-677-682B-7	Sequence 7, Appli
5	1578	100.0	1578	33	US-09-882-694-7	Sequence 7, Appli
6	1578	100.0	1578	33	US-09-882-694A-7	Sequence 7, Appli
7	1578	100.0	1764	17	US-09-351-224E-6	Sequence 6, Appli
8	1578	100.0	1764	26	US-09-677-488A-6	Sequence 6, Appli
9	1578	100.0	1764	26	US-09-677-682A-6	Sequence 6, Appli
10	1578	100.0	1764	26	US-09-677-682B-6	Sequence 6, Appli
11	1578	100.0	1764	33	US-09-882-694-6	Sequence 6, Appli
12	1578	100.0	1764	33	US-09-882-694A-6	Sequence 6, Appli
13	1555.6	98.6	1578	17	US-09-351-224-7	Sequence 7, Appli
14	1555.6	98.6	1578	17	US-09-351-823-7	Sequence 7, Appli
15	1555.6	98.6	1578	26	US-09-677-488-7	Sequence 7, Appli
16	1555.6	98.6	1578	26	US-09-677-682-7	Sequence 7, Appli
17	1555.6	98.6	1764	17	US-09-351-224-6	Sequence 6, Appli
18	1555.6	98.6	1764	17	US-09-351-823-6	Sequence 6, Appli
19	1555.6	98.6	1764	26	US-09-677-488-6	Sequence 6, Appli
20	1555.6	98.6	1764	26	US-09-677-682-6	Sequence 6, Appli
21	106.8	6.8	609	18	US-09-417-507-2341	Sequence 2341, Ap
22	91.8	5.8	1632	18	US-09-404-520-15131	Sequence 15131, A
23	90.2	5.7	837	52	US-60-138-103-7348	Sequence 7348, Ap
24	89.4	5.7	483	18	US-09-417-507-2340	Sequence 2340, Ap
25	86.2	5.5	1466	18	US-09-404-520-19952	Sequence 19952, A
26	86.2	5.5	1504	52	US-60-138-103-5913	Sequence 5913, Ap
27	86.2	5.5	1734	18	US-09-404-520-1908	Sequence 1908, Ap
28	84.2	5.3	807	26	US-09-675-784A-72	Sequence 72, Appl
29	81.2	5.1	864	18	US-09-417-507-15475	Sequence 15475, A
30	79.2	5.0	1980	47	US-60-082-300-10911	Sequence 10911, A
31	78.2	5.0	1981	18	US-09-404-520-20192	Sequence 20192, A

	32	78.2	5.0	2734	18	US-09-404-520-20302	Sequence	20302, A
C	33	78.2	5.0	5837	52	US-60-138-103-7923	Sequence	7923, Ap
C	34	78.2	5.0	7194	18	US-09-404-520-3633	Sequence	3633, Ap
	35	76.6	4.9	549	18	US-09-417-507-2343	Sequence	2343, Ap
	36	75.8	4.8	1194	18	US-09-404-520-19708	Sequence	19708, A
	37	75.8	4.8	1420	18	US-09-404-520-19922	Sequence	19922, A
	38	75.8	4.8	1517	52	US-60-138-103-3579	Sequence	3579, Ap
C	39	75.8	4.8	5225	18	US-09-404-520-4899	Sequence	4899, Ap
	40	73.2	4.6	1059	18	US-09-417-507-17478	Sequence	17478, A
	41	71	4.5	1236	18	US-09-417-507-11427	Sequence	11427, A
C	42	71	4.5	1293	18	US-09-417-507-11429	Sequence	11429, A
	43	69.8	4.4	585	18	US-09-417-507-17798	Sequence	17798, A
C	44	68.2	4.3	696	47	US-60-082-300-5633	Sequence	5633, Ap
	45	68.2	4.3	2602	18	US-09-404-520-20294	Sequence	20294, A

ALIGNMENTS

```

RESULT      1
US-09-351-224E-7
; Sequence 7, Application US/09351224E
; GENERAL INFORMATION:
; APPLICANT: Duwick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; TITLE OF INVENTION: Detoxification
; FILE REFERENCE: 5718-111
; CURRENT APPLICATION NUMBER: US/09/351,224E
; CURRENT FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1578
; TYPE: DNA
; ORGANISM: Exophiala spinifera
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: permealase, fully spliced cDNA
US-09-351-224E-7

```

	Query Match	100.0%;	Score 1578;	DB 17;	Length 1578;
	Best Local Similarity	100.0%;	Pred. NO. 0;		
	Matches 1578;	Conservative	0;	Mismatches	0;
				Indels	Gaps
QY	1	atggactccagacc	aaagtggatactagcgagaaaaggcgggacaaaggcgagacaacgaagaac	60	
Db	1	atggactccagacc	aaagtggatactagcgagaaaaggcgggacaaaggcgagacaacgaagaac	60	
QY	61	acagagacggcg	cgaggtggtgcgtccgagtcctgaacgttcctctggagaagaaa	120	
Db	61	acagagacggcg	cgaggtggtgcgtccgagtcctgaacgttcctctggagaagaaa	120	
QY	121	caatttggcaccat	caccatcgtctccttggcccttctgtgatttgcacagttgggctggt	180	
Db	121	caatttggcaccat	caccatcgtctccttggcccttctgtgatttgcacagttgggctggt	180	
QY	181	atctcaggcagtc	tccagctcgcgccctactagcgggggggcccgtcactctcctttacggc	240	
Db	181	atctcaggcagtc	tccagctcgcgccctactagcgggggggcccgtcactctcctttacggc	240	
QY	241	atcctaatacag	tactctcgtctacatctgcacgtctttctcatttagccgaactgaccagc	300	
Db	241	atcctaatacag	tactctcgtctacatctgcacgtctttctcatttagccgaactgaccagc	300	
QY	301	gtctaccgcact	gcgggtggccaatatcatcttggctcgatcctggcaccacaaatcaatc	360	
Db	301	gtctaccgcact	gcgggtggccaatatcatcttggctcgatcctggcaccacaaatcaatc	360	

Qy 1441 tacacaagtcgattatagcgttgcaactgtctcttgggtgtcttgaactgggtcgtgcat 1500
Db 1441 tacacaagtcgattatagcgttgcaactgtctcttgggtgtcttgaactgggtcgtgcat 1500
Qy 1501 gccaggaagcattatcaggagaccaccacttgagcttgacggagcgggtcgtcggagcagaa 1560
Db 1501 gccaggaagcattatcaggagaccaccacttgagcttgacggagcgggtcgtcggagcagaa 1560
Qy 1561 ttccaagttggccatga 1578
Db 1561 ttccaagttggccatga 1578

RESULT 2
US-09-677-488A-7
; Sequence 7, Application US/09677488A
; GENERAL INFORMATION:
; APPLICANT: DuVick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; TITLE OF INVENTION: Detoxification
; FILE REFERENCE: 35718/204100
; CURRENT APPLICATION NUMBER: US/09/677,488A
; CURRENT FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1578
; TYPE: DNA
; ORGANISM: Exophiala spinifera
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)..(0)
; OTHER INFORMATION: permease, fully spliced cDNA
US-09-677-488A-7

Query Match 100.0%; Score 1578; DB 26; Length 1578;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1578; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atggactccagaccagtcgatacggcgagaaagcgggacaaagcagacaaacgaagaac 60
Db 1 atggactccagaccagtcgatacggcgagaaagcgggacaaagcagacaaacgaagaac 60
Qy 61 acagagacggcgggcggcaggtggtgcgtccgagtcacctgaacgttctcttgagagaagaa 120
Db 61 acagagacggcgggcggcaggtggtgcgtccgagtcacctgaacgttctcttgagagaagaa 120
Qy 121 caatttggcaccatcaccatcgtgtcttggccttctgtgatttgaacagttgggctggt 180
Db 121 caatttggcaccatcaccatcgtgtcttggccttctgtgatttgaacagttgggctggt 180
Qy 181 atctcaggcagtcctcagctcgcctactagcggggggggcccgctcactctcctttacggc 240
Db 181 atctcaggcagtcctcagctcgcctactagcggggggggcccgctcactctcctttacggc 240
Qy 241 atcctaactcagtcactctcgtctacatctgcagcttctctcattagccgaactgaccagc 300
Db 241 atcctaactcagtcactctcgtctacatctgcagcttctctcattagccgaactgaccagc 300
Qy 301 gtctaccgactgcccgttgccaaatatcatcttgcgtcgatccctggcaccacaaatcaatc 360
Db 301 gtctaccgactgcccgttgccaaatatcatcttgcgtcgatccctggcaccacaaatcaatc 360
Qy 361 aatcggagcatttcatacgtgtgcggagctcgttgcgttcttcagatcgctatcgga 420
Db 361 aatcggagcatttcatacgtgtgcggagctcgttgcgttcttcagatcgctatcgga 420

Qy 421 agctcagtgaccatatacctgctcaacagatcccggcgctgatagcgcctatagtcac 480
Db 421 agctcagtgaccatatacctgctcaacagatcccggcgctgatagcgcctatagtcac 480
Qy 481 acatactcccaggattcgtggcatgtcttctcctcatctacgagggagtcgcgtggtggtg 540
Db 481 acatactcccaggattcgtggcatgtcttctcctcatctacgagggagtcgcgtggtggtg 540
Qy 541 ctcttgttcaacttgttgcctgaaagaaaccccttgggttcatgaaatcggtattcggc 600
Db 541 ctcttgttcaacttgttgcctgaaagaaaccccttgggttcatgaaatcggtattcggc 600
Qy 601 ctcaagatcgctctcttctgtagctctcctttatcgccatttctagcgggtccaaacccaaag 660
Db 601 ctcaagatcgctctcttctgtagctctcctttatcgccatttctagcgggtccaaacccaaag 660
Qy 661 gctccaaactcacaggtatgactgcttgagcaactatatactggtggtggtcggagcggctc 720
Db 661 gctccaaactcacaggtatgactgcttgagcaactatatactggtggtggtcggagcggctc 720
Qy 721 tggctcatcctgggcttctgacatcctgcttcatgttcatgttgggtggagcagcaatg 780
Db 721 tggctcatcctgggcttctgacatcctgcttcatgttcatgttgggtggagcagcaatg 780
Qy 781 catctggctgaagaatgcacagatgctgctgtagcgttaccacaaagcagtggtcagtgca 840
Db 781 catctggctgaagaatgcacagatgctgctgtagcgttaccacaaagcagtggtcagtgca 840
Qy 841 atcataaattggcttctgacccgcttctccatatacatacaatgcagttctgtatggaattaca 900
Db 841 atcataaattggcttctgacccgcttctccatatacatacaatgcagttctgtatggaattaca 900
Qy 901 gatctcgactctatttcaagttccgcggctgctatattccattcgagacaatgacgcagtc 960
Db 901 gatctcgactctatttcaagttccgcggctgctatattccattcgagacaatgacgcagtc 960
Qy 961 ctctcgtcgtcagttttgcaacgggtcctctcatgttgcgtatggtggtggtggttcttct 1020
Db 961 ctctcgtcgtcagttttgcaacgggtcctctcatgttgcgtatggtggtggtggttcttct 1020
Qy 1021 gccctcaacgctgtacaaagagactgcgtctcgactcacctggagcttgcggggaacaat 1080
Db 1021 gccctcaacgctgtacaaagagactgcgtctcgactcacctggagcttgcggggaacaat 1080
Qy 1081 gggctggtattttccactcatctcgaacgcatctcccgctggcgaagtctcctgtttgg 1140
Db 1081 gggctggtattttccactcatctcgaacgcatctcccgctggcgaagtctcctgtttgg 1140
Qy 1141 tctctattcgcgacctggggaattcttgcacatgcggatgtatatattctaggttctagc 1200
Db 1141 tctctattcgcgacctggggaattcttgcacatgcggatgtatatattctaggttctagc 1200
Qy 1201 acagctttcaaatgccttgggtcaattccgcgcttctactccagcaactctccttctgact 1260
Db 1201 acagctttcaaatgccttgggtcaattccgcgcttctactccagcaactctccttctgact 1260
Qy 1261 ccaatcgcctactcctctaccacaaagcgagatccaaagtcttgcggagcactcgtgct 1320
Db 1261 ccaatcgcctactcctctaccacaaagcgagatccaaagtcttgcggagcactcgtgct 1320
Qy 1321 ttgtgttaccgctgggaatcgggtttctgtggtcaatgtgtagcgggtgttcttccacgtcc 1380
Db 1321 ttgtgttaccgctgggaatcgggtttctgtggtcaatgtgtagcgggtgttcttccacgtcc 1380
Qy 1381 gtcaccactggttttttcagcttcccactgacctgacctgacctgacctgacctgacctgacct 1440
Db 1381 gtcaccactggttttttcagcttcccactgacctgacctgacctgacctgacctgacctgacct 1440
Qy 1441 tacacaagtcgattataggggttgacattgtcttgggtgtcttgaactgggtcgtgcat 1500
Db 1441 tacacaagtcgattataggggttgacattgtcttgggtgtcttgaactgggtcgtgcat 1500

```
QY 1501 gccaggagcattatcaggagaccccaacttggagcttgacggacgggtcgtcgagcagaa 1560
Db 1501 gccaggagcattatcaggagaccccaacttggagcttgacggacgggtcgtcgagcagaa 1560

QY 1561 ttccaagtgggccatga 1578
Db 1561 ttccaagtgggccatga 1578

RESULT 3
US-09-677-682A-7
; Sequence 7, Application US/09677682A
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; TITLE OF INVENTION: Detoxification
; FILE REFERENCE: 35718/204101
; CURRENT APPLICATION NUMBER: US/09/677,682A
; CURRENT FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1578
; TYPE: DNA
; ORGANISM: Exophiala spinifera
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)..(0)
; OTHER INFORMATION: permease, fully spliced cDNA
US-09-677-682A-7

Query Match 100.0%; Score 1578; DB 26; Length 1578;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1578; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggactccagaccagtggtatcggtcgagaaaggcggagacaggcagacaaacgaagaac 60
Db 1 atggactccagaccagtggtatcggtcgagaaaggcggagacaggcagacaaacgaagaac 60

QY 61 acagagacggcgggcaggtggtgctgcgcagtcctgaacgttcctctgagagagaaa 120
Db 61 acagagacggcgggcaggtggtgctgcgcagtcctgaacgttcctctgagagagaaa 120

QY 121 caattggcaccatcacacatcggtgcttccttgcccttggattgaaacagttgggtggt 180
Db 121 caattggcaccatcacacatcggtgcttccttgcccttggattgaaacagttgggtggt 180

QY 181 atctcaggcagttctccagctcgccctactagcggggggcccgctcactctcctttacggc 240
Db 181 atctcaggcagttctccagctcgccctactagcggggggcccgctcactctcctttacggc 240

QY 241 atcctaactcagttactctcgtctacatctgcatcgctttctcattagccgaactgaccagc 300
Db 241 atcctaactcagttactctcgtctacatctgcatcgctttctcattagccgaactgaccagc 300

QY 301 gtctaccgactgccgggtggcccaataatcatcttgcgtcgatcctctggcaccacaaatcaatc 360
Db 301 gtctaccgactgccgggtggcccaataatcatcttgcgtcgatcctctggcaccacaaatcaatc 360

QY 361 aatcgagcatttctacatcggtgcggaactcggtcggttgccttcattggtcgctatcgga 420
Db 361 aatcgagcatttctacatcggtgcggaactcggtcggttgccttcattggtcgctatcgga 420

QY 421 agctcagtgaccatgatacctgctcaacagatcccgcgctgatacgccctatagtcac 480
Db 421 agctcagtgaccatgatacctgctcaacagatcccgcgctgatacgccctatagtcac 480
```

```
QY 481 acatactccaggattcgtggcatgtcttctctcatctacagggagtcgcgtggtggtg 540
Db 481 acatactccaggattcgtggcatgtcttctctcatctacagggagtcgcgtggtggtg 540

QY 541 ctcttggtcaacttgttgcctgaaagaaaccccttgggttcataaatacggattcggc 600
Db 541 ctcttggtcaacttgttgcctgaaagaaaccccttgggttcataaatacggattcggc 600

QY 601 ctacgactcgtctctcgtgatctcctttatcgccattctagcgggttccaaaccccaag 660
Db 601 ctacgactcgtctctcgtgatctcctttatcgccattctagcgggttccaaaccccaag 660

QY 661 gctccaaactcacaggtatggactgcttggagcaactatactggtggtccgacggcgtc 720
Db 661 gctccaaactcacaggtatggactgcttggagcaactatactggtggtccgacggcgtc 720

QY 721 tgcctcactcctggccttctgacatcctgcttcatgttcatattggttggacgagcaatg 780
Db 721 tgcctcactcctggccttctgacatcctgcttcatgttcatattggttggacgagcaatg 780

QY 781 catctggctgaagaatgcacagatgctgctgcgtacgggtacccaaagcagtggtcagtgca 840
Db 781 catctggctgaagaatgcacagatgctgctgcgtacgggtacccaaagcagtggtcagtgca 840

QY 841 atcataaattggcttctgcacccgcttccatatacaaatcgacgttctgtatggaaattaca 900
Db 841 atcataaattggcttctgcacccgcttccatatacaaatcgacgttctgtatggaaattaca 900

QY 901 gatctcgactctattctaaagtccgcggcttatattccattctgagacaaatgacgcagtc 960
Db 901 gatctcgactctattctaaagtccgcggcttatattccattctgagacaaatgacgcagtc 960

QY 961 ctctggctcgtcagtttgcacgggtcctctcatgtggtggtatcgtggtggtccttcttc 1020
Db 961 ctctggctcgtcagtttgcacgggtcctctcatgtggtggtatcgtggtggtccttcttc 1020

QY 1021 gccctcaacgctgtacaagagactgcgtctcgtcactcacctggagcttgcgcgggacaat 1080
Db 1021 gccctcaacgctgtacaagagactgcgtctcgtcactcacctggagcttgcgcgggacaat 1080

QY 1081 gggctgggtatttccactcatctcgaacgcaattcatcccccgtggcaagtctcctgtttgg 1140
Db 1081 gggctgggtatttccactcatctcgaacgcaattcatcccccgtggcaagtctcctgtttgg 1140

QY 1141 tctctattcgcgacctggggaattcttggccacatgcggatgtatatcttctaggttctagc 1200
Db 1141 tctctattcgcgacctggggaattcttggccacatgcggatgtatatcttctaggttctagc 1200

QY 1201 acagctttcaatgccttgggtcaattccgcgcttgactccagcaactctccttctcgtatc 1260
Db 1201 acagctttcaatgccttgggtcaattccgcgcttgactccagcaactctccttctcgtatc 1260

QY 1261 ccaatcgccctactcctctacccaaagcgagatccaaagtcttcttgcgagcactcgtgct 1320
Db 1261 ccaatcgccctactcctctacccaaagcgagatccaaagtcttcttgcgagcactcgtgct 1320

QY 1321 tttgtgttacccggtggaatcggtttctcgttcaatgtgctagcgggtggtcttctcagtc 1380
Db 1321 tttgtgttacccggtggaatcggtttctcgttcaatgtgctagcgggtggtcttctcagtc 1380

QY 1381 gtccaccactgtgttttcaagcttcccaactgacgtggtcactacggcggtcgaacccatgaat 1440
Db 1381 gtccaccactgtgttttcaagcttcccaactgacgtggtcactacggcggtcgaacccatgaat 1440

QY 1441 tacacaagtgcgattatagcgttgcacttgcctcttgggtgttgaactgggtcgtgcac 1500
Db 1441 tacacaagtgcgattatagcgttgcacttgcctcttgggtgttgaactgggtcgtgcac 1500

QY 1501 gccagggaagcattatcaggagcccccacttggagcttgcagggcggtcgtcggagcagaa 1560
Db 1501 gccagggaagcattatcaggagcccccacttggagcttgcagggcggtcgtcggagcagaa 1560
```



```
RESULT 5
US-09-882-694-7
; Sequence 7, Application US/09882694
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; TITLE OF INVENTION: Detoxification
; FILE REFERENCE: 35718/208255
; CURRENT APPLICATION NUMBER: US/09/882,694
; CURRENT FILING DATE: 2001-06-12
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1578
; TYPE: DNA
; ORGANISM: Exophiala spinifera
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: permease, fully spliced cDNA
US-09-882-694-7
```

```
Query Match 100.0%; Score 1578; DB 33; Length 1578;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1578; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 atggactccagaccaggaagtggatagcggcgagaaagcggggacaaaggcgagacaacgaagaac 60
Db 1 atggactccagaccaggaagtggatagcggcgagaaagcggggacaaaggcgagacaacgaagaac 60

QY 61 acagagacggcgggcgaggtggtgcgtccgagtcgcagtcctgaacgttccctctgagagagaaa 120
Db 61 acagagacggcgggcgaggtggtgcgtccgagtcgcagtcctgaacgttccctctgagagagaaa 120

QY 121 caatttggccaccatcaccatcgtgtcccttggcccttctgtgatttgcacagttgggctgggt 180
Db 121 caatttggccaccatcaccatcgtgtcccttggcccttctgtgatttgcacagttgggctgggt 180

QY 181 atctcaggcagtcctccagtcgcgcctactagcggggggggcccgctacatcctcttaccggc 240
Db 181 atctcaggcagtcctccagtcgcgcctactagcggggggggcccgctacatcctcttaccggc 240

QY 241 atcctaatacagtaactctcgtctacatctgcagtcgcttctcattagcgaactgaccagc 300
Db 241 atcctaatacagtaactctcgtctacatctgcagtcgcttctcattagcgaactgaccagc 300

QY 301 gtctacccgactgccgggtggccaatatcatcttgcgtcgatccctggcaccacaaatcaatc 360
Db 301 gtctacccgactgccgggtggccaatatcatcttgcgtcgatccctggcaccacaaatcaatc 360

QY 361 aatcggagcatttcatacagtcgtgcggactcgttcgttctcattagcgaactgctatcgga 420
Db 361 aatcggagcatttcatacagtcgtgcggactcgttcgttctcattagcgaactgctatcgga 420

QY 421 agctcagtgaccatgatacctgtctcaacagatcccgcgctgatagcgcgcctatagtcac 480
Db 421 agctcagtgaccatgatacctgtctcaacagatcccgcgctgatagcgcgcctatagtcac 480

QY 481 acatactccaggattcgtggcatgtcttccctcatctacagaggggagtcgctgggtgggtg 540
Db 481 acatactccaggattcgtggcatgtcttccctcatctacagaggggagtcgctgggtgggtg 540

QY 541 ctctgttcaactgtttgcccctgaaagaaaccccttgggttcatgaaatcggattccggc 600
Db 541 ctctgttcaactgtttgcccctgaaagaaaccccttgggttcatgaaatcggattccggc 600
```

```
QY 601 ctacagatcgctctcttctcgtgatctcctctttatcgccattcttagcgcggtcccaacccccaaag 660
Db 601 ctacagatcgctctcttctcgtgatctcctctttatcgccattcttagcgcggtcccaacccccaaag 660

QY 661 gctccaaaactcacaggtatggactgcttggagcaactatatactggtggtccgacggcgctc 720
Db 661 gctccaaaactcacaggtatggactgcttggagcaactatatactggtggtccgacggcgctc 720

QY 721 tgccttcactcctgggccccttccgacatcctgcttcatgttcatgttgccttggagcagcaatg 780
Db 721 tgccttcactcctgggccccttccgacatcctgcttcatgttcatgttgccttggagcagcaatg 780

QY 781 catctggctgaagaatgcacagatgctgctcgtacggtacccaaagcagtggtcagtgca 840
Db 781 catctggctgaagaatgcacagatgctgctcgtacggtacccaaagcagtggtcagtgca 840

QY 841 atcataattggccttctgcaccgcttcccatatacaaatcgagttctgtatggaattaca 900
Db 841 atcataattggccttctgcaccgcttcccatatacaaatcgagttctgtatggaattaca 900

QY 901 gatctcgactctattctaaagtctccgggcttatctcagactcacctggagccttggcccgggacaat 960
Db 901 gatctcgactctattctaaagtctccgggcttatctcagactcacctggagccttggcccgggacaat 960

QY 961 cttcgggtcgctcagtttttgcaacgggtcctctcctcgtatggtgggtatcgtgagtccttcttc 1020
Db 961 cttcgggtcgctcagtttttgcaacgggtcctctcctcgtatggtgggtatcgtgagtccttcttc 1020

QY 1021 gccctcaacgctgtacaaagagactgctcgtcgcactcacctggagccttggcccgggacaat 1080
Db 1021 gccctcaacgctgtacaaagagactgctcgtcgcactcacctggagccttggcccgggacaat 1080

QY 1081 gggctgggtattttccactcatctcgaacgcattcattcccccgctggcaagtctcctgtttgg 1140
Db 1081 gggctgggtattttccactcatctcgaacgcattcattcccccgctggcaagtctcctgtttgg 1140

QY 1141 tctctattcgcgacctgggggaattctggccacatgctgggatgtatatattcttaggttctagc 1200
Db 1141 tctctattcgcgacctgggggaattctggccacatgctgggatgtatatattcttaggttctagc 1200

QY 1201 acagctttcaatgcttggtaattctggccacatgctgggatgtatatattcttaggttctagc 1260
Db 1201 acagctttcaatgcttggtaattctggccacatgctgggatgtatatattcttaggttctagc 1260

QY 1261 ccaatcgccctactcctctacaaaagcgagatccaaagtcttgcgagcactcgtgct 1320
Db 1261 ccaatcgccctactcctctacaaaagcgagatccaaagtcttgcgagcactcgtgct 1320

QY 1321 tttgtgtaccgctggtgaatcggtttcttggtcaatgtgctagcgtggtgttcttcacgctc 1380
Db 1321 tttgtgtaccgctggtgaatcggtttcttggtcaatgtgctagcgtggtgttcttcacgctc 1380

QY 1381 gtcaccactgtgttttccagcttcccaactgaccgtgcctacggtgcgtcaacccatgaat 1440
Db 1381 gtcaccactgtgttttccagcttcccaactgaccgtgcctacggtgcgtcaacccatgaat 1440

QY 1441 tacacaagtgcgattataggggttgcaacttgctcttggtgtcttgaactgggtcgtgcat 1500
Db 1441 tacacaagtgcgattataggggttgcaacttgctcttggtgtcttgaactgggtcgtgcat 1500

QY 1501 gccaggaagcattatcacgggggacccccacttggagcttgacggacgggtgctcggagcagaa 1560
Db 1501 gccaggaagcattatcacgggggacccccacttggagcttgacggacgggtgctcggagcagaa 1560

QY 1561 ttccaagtggggccatga 1578
Db 1561 ttccaagtggggccatga 1578
```

```
RESULT 6
US-09-882-694A-7
; Sequence 7, Application US/09882694A
```

; GENERAL INFORMATION:
; APPLICANT: Duvick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; TITLE OF INVENTION: Detoxification
; FILE REFERENCE: 5718-111
; CURRENT APPLICATION NUMBER: US/09/882,694A
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1578
; TYPE: DNA
; ORGANISM: Exophiala spinifera
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)..(0)
; OTHER INFORMATION: permease, fully spliced cdna
US-09-882-694A-7

Query Match 100.0%; Score 1578; DB 33; Length 1578;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1578; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggactccagaccagtggtgatacggcgagaaaggcggaacaggcgagacacgaagaac 60
Db 1 atggactccagaccagtggtgatacggcgagaaaggcggaacaggcgagacacgaagaac 60
QY 61 acagagacggcgggcagggtggtgcgtccgagtcacctgaacgttcctctgagagaaga 120
Db 61 acagagacggcgggcagggtggtgcgtccgagtcacctgaacgttcctctgagagaaga 120
QY 121 caatttggcaccatcaccatcgctgtcccttggcctttgtgatttgcacacgttgggtggt 180
Db 121 caatttggcaccatcaccatcgctgtcccttggcctttgtgatttgcacacgttgggtggt 180
QY 181 atctcaggcagtcctcagctgcctactagcggggggcgctcactcctccttactcggc 240
Db 181 atctcaggcagtcctcagctgcctactagcggggggcgctcactcctccttactcggc 240
QY 241 atcctaatacagtaactcgtctacatctgcacgtcttctcattagccgaactgaccagc 300
Db 241 atcctaatacagtaactcgtctacatctgcacgtcttctcattagccgaactgaccagc 300
QY 301 gtctacccgactgcgggtggcgaataatcatttgcgtcgatcctggcaccacaaatac 360
Db 301 gtctacccgactgcgggtggcgaataatcatttgcgtcgatcctggcaccacaaatac 360
QY 361 aatcggagcatttcaatcgtgtgcggactcgtgtcgttgcattcattggtatcgctatcgga 420
Db 361 aatcggagcatttcaatcgtgtgcggactcgtgtcgttgcattcattggtatcgctatcgga 420
QY 421 agctcagtgaccatgatacctgctcaacagatcccggcgtgtagatagccctatagtcac 480
Db 421 agctcagtgaccatgatacctgctcaacagatcccggcgtgtagatagccctatagtcac 480
QY 481 acatactcccaggattcgtggtcgtgtcctcctcattcctcagggaggtcgcgtggtggtg 540
Db 481 acatactcccaggattcgtggtcgtgtcctcctcattcctcagggaggtcgcgtggtggtg 540
QY 541 ctcttggtcaactgttttgcctgaaagaaaccccttgggttccatgaaatcggattcggc 600
Db 541 ctcttggtcaactgttttgcctgaaagaaaccccttgggttccatgaaatcggattcggc 600
QY 601 ctcaagatcgcctctctctgtgatctccttctatcgccattctagcgggtcccaacccaag 660
Db 601 ctcaagatcgcctctctctgtgatctccttctatcgccattctagcgggtcccaacccaag 660

QY 661 gctccaaactcacaggtatggactgttggagcaactatactggctgggtccgagcgcgtc 720
Db 661 gctccaaactcacaggtatggactgttggagcaactatactggctgggtccgagcgcgtc 720
QY 721 tgcttcactcctggccttctcgacatcctgcttcatgttcatgttgccttggaacgaagaatg 780
Db 721 tgcttcactcctggccttctcgacatcctgcttcatgttcatgttgccttggaacgaagaatg 780
QY 781 catctggtggaagaatgcacagatgctgctcgtacgggtacccaaagcagtggtcagtgca 840
Db 781 catctggtggaagaatgcacagatgctgctcgtacgggtacccaaagcagtggtcagtgca 840
QY 841 atcataaattggcttctgcacgccttctccatatatacaaatcgcagttctgtatggaattaca 900
Db 841 atcataaattggcttctgcacgccttctccatatatacaaatcgcagttctgtatggaattaca 900
QY 901 gatctcgactctattctaagtctccgcgggtctatatctccattcggagacaatgacgcagtc 960
Db 901 gatctcgactctattctaagtctccgcgggtctatatctccattcggagacaatgacgcagtc 960
QY 961 cttcgggtcgtcagtttgcacgggtcctctcatgtggcggtatcgtgatggccttcttc 1020
Db 961 cttcgggtcgtcagtttgcacgggtcctctcatgtggcggtatcgtgatggccttcttc 1020
QY 1021 gccctcaacgctgtacaagagactgctcgcactcaacctggagcttgcggggaacaat 1080
Db 1021 gccctcaacgctgtacaagagactgctcgcactcaacctggagcttgcggggaacaat 1080
QY 1081 gggctggtattttccactcatctcgaaacgcatctcccgctggaagtctcctgttgg 1140
Db 1081 gggctggtattttccactcatctcgaaacgcatctcccgctggaagtctcctgttgg 1140
QY 1141 tctctattcgcgacctggggaattcttggccacatgctggtatctatttctaggttctagc 1200
Db 1141 tctctattcgcgacctggggaattcttggccacatgctggtatctatttctaggttctagc 1200
QY 1201 acagctttcaaatgccttgggtcaattcgcgcgttcttactccagcaactcctctcctgac 1260
Db 1201 acagctttcaaatgccttgggtcaattcgcgcgttcttactccagcaactcctctcctgac 1260
QY 1261 ccaatgcgcctactcctctaccacaaagcgagatccaaagtcttctgcgagcactcgtgct 1320
Db 1261 ccaatgcgcctactcctctaccacaaagcgagatccaaagtcttctgcgagcactcgtgct 1320
QY 1321 tttgtgttacccgctggaatcgggttcttgggtcaatgtgctagcgttggcttccagtc 1380
Db 1321 tttgtgttacccgctggaatcgggttcttgggtcaatgtgctagcgttggcttccagtc 1380
QY 1381 gtacccactgtgttttccagcttcccaactgacctgacctgacctgacctgacctgacct 1440
Db 1381 gtacccactgtgttttccagcttcccaactgacctgacctgacctgacctgacctgacct 1440
QY 1441 tacacaaagtgcgattatagggcgttgcacttgccttcttgggtcttgaactgggtcgtgcat 1500
Db 1441 tacacaaagtgcgattatagggcgttgcacttgccttcttgggtcttgaactgggtcgtgcat 1500
QY 1501 gccaggaagcattatcagggaccccaacttggagcttgacggagcgggtcgtcggagcagaa 1560
Db 1501 gccaggaagcattatcagggaccccaacttggagcttgacggagcgggtcgtcggagcagaa 1560
QY 1561 ttccaagttgggccaatga 1578
Db 1561 ttccaagttgggccaatga 1578

RESULT 7
US-09-351-224E-6
; Sequence 6, Application US/09351224E
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob

; CURRENT FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 1764
; TYPE: DNA
; ORGANISM: Exophiala spinifera
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: permealase, partially spliced cDNA
US-09-677-488A-6

Query Match 100.0%; Score 1578; DB 26; Length 1764;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1578; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	atggactccagacaaagtggatatacggcgagaaaggcgggacaaaggcagacaacgaagaac	60
Db	5	atggactccagacaaagtggatatacggcgagaaaggcgggacaaaggcagacaacgaagaac	64
QY	61	acagagacggcgcgaggggtggtgcgtccgagtcgcctgaacgttccctggagaaagaa	120
Db	65	acagagacggcgcgaggggtggtgcgtccgagtcgcctgaacgttccctggagaaagaa	124
QY	121	caatttggcaccatcaccatcgtgtcccttggcctttgtgatttgcaacagttggctggt	180
Db	125	caatttggcaccatcaccatcgtgtcccttggcctttgtgatttgcaacagttggctggt	184
QY	181	atctcaggcagtcctccagtcgccctactagcggggggcccgctcactctcctttacggc	240
Db	185	atctcaggcagtcctccagtcgccctactagcggggggcccgctcactctcctttacggc	244
QY	241	atcctaatacgtactctcgtctacatctgcacgtcttctcatttagccgaactgaccagc	300
Db	245	atcctaatacgtactctcgtctacatctgcacgtcttctcatttagccgaactgaccagc	304
QY	301	gtctacccgactccgggtggccaaatatcatttgcgtcgatccctggcaccacaaatcaatc	360
Db	305	gtctacccgactccgggtggccaaatatcatttgcgtcgatccctggcaccacaaatcaatc	364
QY	361	aatcggagcatttcatacgtgtcgggactcgtgtcgttgccttcatggatcgctatcgga	420
Db	365	aatcggagcatttcatacgtgtcgggactcgtgtcgttgccttcatggatcgctatcgga	424
QY	421	agctcagtgaccatgatacctgctcaacagatccccggcgctgatagccgcctatagtcac	480
Db	425	agctcagtgaccatgatacctgctcaacagatccccggcgctgatagccgcctatagtcac	484
QY	481	acatactccaggattcgtggcatgtcttccatctacagaggagtcggcgtggtggtg	540
Db	485	acatactccaggattcgtggcatgtcttccatctacagaggagtcggcgtggtggtg	544
QY	541	ctcttggtcaacttggttgcctgaaagaaaccccttgggttcattgaaatcggattcggc	600
Db	545	ctcttggtcaacttggttgcctgaaagaaaccccttgggttcattgaaatcggattcggc	604
QY	601	ctcagcatgctctctctcgtgatctccttctatcgccatttagcgcgttccaccccaag	660
Db	605	ctcagcatgctctctctcgtgatctccttctatcgccatttagcgcgttccaccccaag	664
QY	661	gctccaaactcacaggataggactgcttgaggcaactatactggtggtcgcagcgcgtc	720
Db	665	gctccaaactcacaggataggactgcttgaggcaactatactggtggtcgcagcgcgtc	724
QY	721	tgcttcatcctgggccccttccgacatccttctcattgatttgccttggcagcagcaatg	780
Db	725	tgcttcatcctgggccccttccgacatccttctcattgatttgccttggcagcagcaatg	784
QY	781	catctggctgaagaatgcacagatgctgctcgtacggtaccacaaagcagtggtcagtgca	840

Db	785	catctggctgaagaatgcacagatgctgctcgtacggtaccacaaagcagtggtcagtgca	844
QY	841	atcataaattggcttctgcacgcctttccatatatacaaatcgaggttctgtatgaattaca	900
Db	845	atcataaattggcttctgcacgcctttccatatatacaaatcgaggttctgtatgaattaca	904
QY	901	gatctcgactctattctaaagtccgcggctctatatattccattcgagacaatgacgcagtc	960
Db	905	gatctcgactctattctaaagtccgcggctctatatattccattcgagacaatgacgcagtc	964
QY	961	cttcgggtcgctcagttttgcaacgggtccctctcatgtggcggtatcgtgatggccttcttc	1020
Db	965	cttcgggtcgctcagttttgcaacgggtccctctcatgtggcggtatcgtgatggccttcttc	1024
QY	1021	gcctcaaacgctgtacaaagagactgcgtctcgactcaactcaactggagctttgcccggaaca	1080
Db	1025	gcctcaaacgctgtacaaagagactgcgtctcgactcaactcaactggagctttgcccggaaca	1084
QY	1081	gggctggtattttccactcattctcgaaacgcattcatccccgcgtggcaagtctcctgtttgg	1140
Db	1085	gggctggtattttccactcattctcgaaacgcattcatccccgcgtggcaagtctcctgtttgg	1144
QY	1141	tctctattcgcgaacctgggggaattctggccacatgcggatgtatatattcttaggttttagc	1200
Db	1145	tctctattcgcgaacctgggggaattctggccacatgcggatgtatatattcttaggttttagc	1204
QY	1201	acaqctttcaatgcttgggtcaaatccgcggttctgtactccagaaactctccttctcgtac	1260
Db	1205	acagctttcaatgcttgggtcaaatccgcggttctgtactccagaaactctccttctcgtac	1264
QY	1261	ccaatcgccctactcctctacaaaagcgagatcccaaatcttctgcgagcactcgtgct	1320
Db	1265	ccaatcgccctactcctctacaaaagcgagatcccaaatcttctgcgagcactcgtgct	1324
QY	1321	tttgtgttacgcgtggaatcgggtttctgtgccaatgtgctagcgggtggtcttccacgtcc	1380
Db	1325	tttgtgttacgcgtggaatcgggtttctgtgccaatgtgctagcgggtggtcttccacgtcc	1384
QY	1381	gtcaccaactgttttccagcttcccaactgacgtgcctacgcgcgcgtcaacacatgaat	1440
Db	1385	gtcaccaactgttttccagcttcccaactgacgtgcctacgcgcgcgtcaacacatgaat	1444
QY	1441	tacacaaagtgcgattataggcgttgcaacttgccttgccttgcctgaactgggtcgtgcat	1500
Db	1445	tacacaaagtgcgattataggcgttgcaacttgccttgccttgcctgaactgggtcgtgcat	1504
QY	1501	gccagggaagcattatcaggagaccccaacttggagcttgacgagcgggtcgtcgagagcagaa	1560
Db	1505	gccagggaagcattatcaggagaccccaacttggagcttgacgagcgggtcgtcgagagcagaa	1564
QY	1561	tttcaagttgggccaatga	1578
Db	1565	tttcaagttgggccaatga	1582

RESULT 9
US-09-677-682A-6
; Sequence 6, Application US/09677682A
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; TITLE OF INVENTION: Detoxification
; FILE REFERENCE: '35718/204101'
; CURRENT APPLICATION NUMBER: US/09/677,682A
; CURRENT FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11


```
; SOFTWARE: Fast5EQ for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 1764
; TYPE: DNA
; ORGANISM: Exophiala spinifera
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: permealase, partially spliced cDNA
US-09-677-682A-6

Query Match      100.0%; Score 1578; DB 26; Length 1764;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1578; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggactccagaccagaagtggatcacggcgagaaaggcgggacaaaggcagacaaacgaagaac 60
Db 5 atggactccagaccagaagtggatcacggcgagaaaggcgggacaaaggcagacaaacgaagaac 64

QY 61 acagagacggcgggcgaggtgggtgcgtccagagtcacctgaacgttcctctggaagaagaa 120
Db 65 acagagacggcgggcgaggtgggtgcgtccagagtcacctgaacgttcctctggaagaagaa 124

QY 121 caatttggcaccatcacccatcggtgctcttggccctttgtgatttgcacagttgggtggt 180
Db 125 caatttggcaccatcacccatcggtgctcttggccctttgtgatttgcacagttgggtggt 184

QY 181 atctcaggcagctctccagctcgcctactagcggggggggcccgctcactctcctttacggc 240
Db 185 atctcaggcagctctccagctcgcctactagcggggggggcccgctcactctcctttacggc 244

QY 241 atcctaatacgtactctcgtctacatctgcacgtcttctcattagcgaactgaccagc 300
Db 245 atcctaatacgtactctcgtctacatctgcacgtcttctcattagcgaactgaccagc 304

QY 301 gtctaccgactgcgggtggccaaatatcatatttgcgtcgatcctggcaccacaaatacaatc 360
Db 305 gtctaccgactgcgggtggccaaatatcatatttgcgtcgatcctggcaccacaaatacaatc 364

QY 361 aatcggagcatttcatacgtgtgcggactcgtgctgctgtcttccatggatcgctatcgga 420
Db 365 aatcggagcatttcatacgtgtgcggactcgtgctgctgtcttccatggatcgctatcgga 424

QY 421 agctcagtgaccatgataacctgctcaacagatcccgcgctgatagcgcctatagtcac 480
Db 425 agctcagtgaccatgataacctgctcaacagatcccgcgctgatagcgcctatagtcac 484

QY 481 acatactcccaggattcgtggcatgtcttctcctcatctacgagggagtcgcgtggtggtg 540
Db 485 acatactcccaggattcgtggcatgtcttctcctcatctacgagggagtcgcgtggtggtg 544

QY 541 ctcttgtcacaactgttttgcctgaaaagaacccttgggttccatgaaatcggattcggc 600
Db 545 ctcttgttcaactgttttgcctgaaaagaacccttgggttccatgaaatcggattcggc 604

QY 601 ctacgatcgtctctctcgtgatctccttcttatcgccattctagcgggtccaaaccccaag 660
Db 605 ctacgatcgtctctcgtgatctccttcttatcgccattctagcgggtccaaaccccaag 664

QY 661 gctccaaactcacagggtatggactgcttgaggaactatactaggctgggtccgacggcgctc 720
Db 665 gctccaaactcacagggtatggactgcttgaggaactatactaggctgggtccgacggcgctc 724

QY 721 tgccttcacctggcccttttcgacatcctgcttcattgttccatggcttggacgcagcaatg 780
Db 725 tgccttcacctggcccttttcgacatcctgcttcattgttccatggcttggacgcagcaatg 784

QY 781 catctggtgaagaatgcacagatgctgctcgtagcgttacccaaagcagtggttcagtgca 840
Db 785 catctggtgaagaatgcacagatgctgctcgtagcgttacccaaagcagtggttcagtgca 844

QY 841 atcataaattggtctctgcacccgccttttccatatatacaaatcgcagttctgtatggaattaca 900
```

```
Db 845 atcataaattggtctctgcacccgcctttccatatatacaaatcgcagttctgtatggaattaca 904

QY 901 gatctcgactctattctaaagtccggcggcgttatattccatttcgagagacaatgacgcagtc 960
Db 905 gatctcgactctattctaaagtccggcggcgttatattccatttcgagagacaatgacgcagtc 964

QY 961 cttegggtcgctcagtttttgcacagggtcctctctcatgttggcgggtatcgtgatggccttcttc 1020
Db 965 cttegggtcgctcagtttttgcacagggtcctctctcatgttggcgggtatcgtgatggccttcttc 1024

QY 1021 gccctcaacgctgtacaagagagactgcgtctcgcactcaccttgagccttgcgggacaat 1080
Db 1025 gccctcaacgctgtacaagagagactgcgtctcgcactcaccttgagccttgcgggacaat 1084

QY 1081 gggctggtattttccactcatctcgaacgcattccatcccccgcgtggcaagttcctgtttgg 1140
Db 1085 gggctggtattttccactcatctcgaacgcattccatcccccgcgtggcaagttcctgtttgg 1144

QY 1141 tctctattcgcgcacctggggaattcttgccacacatcgcgatgtatatattctaggttcttagc 1200
Db 1145 tctctattcgcgcacctggggaattcttgccacacatcgcgatgtatatattctaggttcttagc 1204

QY 1201 acagctttcaatgccttgggtcaattccgcggttgactccagcaactcctcctcctgac 1260
Db 1205 acagctttcaatgccttgggtcaattccgcggttgactccagcaactcctcctcctgac 1264

QY 1261 ccaatgcctcactcctctacaaaagcgagatccaaaagttcttgcgcgagcactcgtgct 1320
Db 1265 ccaatgcctcactcctctacaaaagcgagatccaaaagttcttgcgcgagcactcgtgct 1324

QY 1321 tttgtgttaccgcgtggaaatcgggtttcttgggtcaatgtgctagcgggtggttcttcacgtcc 1380
Db 1325 tttgtgttaccgcgtggaaatcgggtttcttgggtcaatgtgctagcgggtggttcttcacgtcc 1384

QY 1381 gtcacacactgtgtttttcagcttcccaactgacctgacctgacctgacctgacctgacct 1440
Db 1385 gtcacacactgtgtttttcagcttcccaactgacctgacctgacctgacctgacctgacct 1444

QY 1441 tacacaagtgcgattatataggcgttgcaacttgcacttgccttgccttgccttgccttgcct 1500
Db 1445 tacacaagtgcgattatataggcgttgcaacttgcacttgccttgccttgccttgccttgcct 1504

QY 1501 gccaggaaagcattatcacgggaccccaacttggagcttgacggacgggtcgtcggagcagaa 1560
Db 1505 gccaggaaagcattatcacgggaccccaacttggagcttgacggacgggtcgtcggagcagaa 1564

QY 1561 ttccaagtggggccatga 1578
Db 1565 ttccaagtggggccatga 1582

RESULT 10
US-09-677-682B-6
; Sequence 6, Application US/09677682B
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; TITLE OF INVENTION: Detoxification
; FILE REFERENCE: 35718/204101
; CURRENT APPLICATION NUMBER: US/09/677,682B
; CURRENT FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 1764
; TYPE: DNA
```

; ORGANISM: Exophiala spinifera
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: permease, partially spliced cDNA
US-09-677-682B-6

Query Match 100.0%; Score 1578; DB 26; Length 1764;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1578; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggactccagaccacaaagtggatcacgagcagagaaagcgaggacaaagcgagacaaagaaac 60
Db 5 atggactccagaccacaaagtggatcacgagcagagaaagcgaggacaaagcgagacaaagaaac 64
QY 61 acagagacggcgcgaggtggtgcgtcccgagtcacctgaacgttcccttgaggagaa 120
Db 65 acagagacggcgcgaggtggtgcgtcccgagtcacctgaacgttcccttgaggagaa 124
QY 121 caatttggcaccatcaccatcgtgtcccttggtggtgatttgcaacagttgggtcgt 180
Db 125 caatttggcaccatcaccatcgtgtcccttggtggtgatttgcaacagttgggtcgt 184
QY 181 atctcaggcagtcctccagctgcctcactagcggggggggcccgctcactcctttacggc 240
Db 185 atctcaggcagtcctccagctgcctcactagcggggggggcccgctcactcctttacggc 244
QY 241 atcctaatacgtactctcgtctacatctgcctcttctcatlagccgaactgacacgc 300
Db 245 atcctaatacgtactctcgtctacatctgcctcttctcatlagccgaactgacacgc 304
QY 301 gtctacccgactgcgggtggcacaatatcatcttgcgtcgatccctggcaccacaaatcaatc 360
Db 305 gtctacccgactgcgggtggcacaatatcatcttgcgtcgatccctggcaccacaaatcaatc 364
QY 361 aatcggagcatttccatacgtgtcgggactcgtgtcgttgccttcacgtatcgctacgga 420
Db 365 aatcggagcatttccatacgtgtcgggactcgtgtcgttgccttcacgtatcgctacgga 424
QY 421 agctcagtgaccatgatacctcgtcctcaacagatcccggcgctgatacgcgcctatagtcac 480
Db 425 agctcagtgaccatgatacctcgtcctcaacagatcccggcgctgatacgcgcctatagtcac 484
QY 481 acatactccaggattcgtggcatgtcttccctacatctacgagggagtcgctggtggtg 540
Db 485 acatactccaggattcgtggcatgtcttccctacatctacgagggagtcgctggtggtg 544
QY 541 ctcttgttcaactgtttgcctgaaagaaaccccttgggttcataaatacggattcggc 600
Db 545 ctcttgttcaactgtttgcctgaaagaaaccccttgggttcataaatacggattcggc 604
QY 601 ctacagatcgtctcttctcgtgatctcctttatcgccattctagcgggtcccaaccccaag 660
Db 605 ctacagatcgtctctcgtgatctcctttatcgccattctagcgggtcccaaccccaag 664
QY 661 gctccaaactcaacagggtatggactgcttggagcaactatactggttccgacggcgtc 720
Db 665 gctccaaactcaacagggtatggactgcttggagcaactatactggttccgacggcgtc 724
QY 721 tgettccatcctggcctttcgacatcctgcttccatgttccatgttggttgacgcagcaatg 780
Db 725 tgettccatcctggcctttcgacatcctgcttccatgttccatgttggttgacgcagcaatg 784
QY 781 catctggctgaagaatgcacagatgctgctcgatcgggtaccacaaagcagtggtcagtgca 840
Db 785 catctggctgaagaatgcacagatgctgctcgatcgggtaccacaaagcagtggtcagtgca 844
QY 841 atcataatitggcttctgcaccgcctttccatatatacaatcgcagttctgtatggaaattaca 900
Db 845 atcataatitggcttctgcaccgcctttccatatatacaatcgcagttctgtatggaaattaca 904
QY 901 gatctcgactctattctaaagtctccggcgtctatatctccattccgagacaaatgacgcagtc 960

Db 905 gatctcgactctattctaaagttccgcggctatatctccattcgagacaaatgacgcagtc 964
QY 961 ctctcggtcgtcagttttgcaacggttcctctcatgttggggttatcgtgatggccttcttc 1020
Db 965 ctctcggtcgtcagttttgcaacggttcctctcatgttggggttatcgtgatggccttcttc 1024
QY 1021 gccctcaacgctgtacaagagactgcgtctcgactcacctggagctttgcccgggagacaa 1080
Db 1025 gccctcaacgctgtacaagagactgcgtctcgactcacctggagctttgcccgggagacaa 1084
QY 1081 gggtcgttatcttccactcatctcgaacgattccatcccgctgggcaagtctcctgttgg 1140
Db 1085 gggtcgttatcttccactcatctcgaacgattccatcccgctgggcaagtctcctgttgg 1144
QY 1141 tctctattcgcgacacctggggaattcttgccacatgcggatgtatatattcttaggttctagc 1200
Db 1145 tctctattcgcgacacctggggaattcttgccacatgcggatgtatatattcttaggttctagc 1204
QY 1201 acagctttcaatgccttgggtcaattccgctcggttgcctccagcaactcctctcctgac 1260
Db 1205 acagctttcaatgccttgggtcaattccgctcggttgcctccagcaactcctctcctgac 1264
QY 1261 ccaatcgccctactcctctaccacaaagcgagatccaaagtcttgcgagcactcgtgct 1320
Db 1265 ccaatcgccctactcctctaccacaaagcgagatccaaagtcttgcgagcactcgtgct 1324
QY 1321 ttgtgttaccgctgggaatcggtttctgtgtaaatgtgctagcgggtggttcttcaogtcc 1380
Db 1325 ttgtgttaccgctgggaatcggtttctgtgtaaatgtgctagcgggtggttcttcaogtcc 1384
QY 1381 gtcacacactgtgtttttcagcttccacactgacctgacctgacctgacctgacctgacct 1440
Db 1385 gtcacacactgtgtttttcagcttccacactgacctgacctgacctgacctgacctgacct 1444
QY 1441 tacacaagtgcgattataggcgttgcacttgccttgccttgccttgccttgccttgcct 1500
Db 1445 tacacaagtgcgattataggcgttgcacttgccttgccttgccttgccttgccttgcct 1504
QY 1501 gccagggaagcattatcagggaacccaccttggagcttgacggagcgggtcgtcgagcagaa 1560
Db 1505 gccagggaagcattatcagggaacccaccttggagcttgacggagcgggtcgtcgagcagaa 1564
QY 1561 ttccaagtgtgggccatga 1578
Db 1565 ttccaagtgtgggccatga 1582

RESULT 11
US-09-882-694-6
; Sequence 6, Application US/09882694
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; TITLE OF INVENTION: Detoxification
; FILE REFERENCE: 35718/208255
; CURRENT APPLICATION NUMBER: US/09/882,694
; CURRENT FILING DATE: 2001-06-12
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 1764
; TYPE: DNA
; ORGANISM: Exophiala spinifera
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)

Query Match 100.0%; Score 1578; DB 33; Length 1764; Best Local Similarity 100.0%; Pred. No. 0; Matches 1578; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	atggactccagacc	aaagtggatacggcgagaaaaggcggaacaaggcagacaacgaagaac	60					
Db	5	atggactccagacc	aaagtggatacggcgagaaaaggcggaacaaggcagacaacgaagaac	64					
QY	61	acagagacggcg	ggcgaggtggtgcgtccgagtcctgaacgttccctctggagaagaaa	120					
Db	65	acagagacggcg	ggcgaggtggtgcgtccgagtcctgaacgttccctctggagaagaaa	124					
QY	121	caatttggcacc	atcaccatcggttcccttggcccttggatttggcaacagttggctggt	180					
Db	125	caatttggcacc	atcaccatcggttcccttggcccttggatttggcaacagttggctggt	184					
QY	181	atctcaggcag	tctccagctcgccctactagcggggggcccgctcactctcctttacggc	240					
Db	185	atctcaggcag	tctccagctcgccctactagcggggggcccgctcactctcctttacggc	244					
QY	241	atcctaatac	gactctctcgtctacatctgcacgtcttctcattagccgaactgaccgc	300					
Db	245	atcctaatac	gactctctcgtctacatctgcacgtcttctcattagccgaactgaccgc	304					
QY	301	gtctaccgac	tgcgggtggccaatatcatcttggctcgatcctggcaccaaaatcaatc	360					
Db	305	gtctaccgac	tgcgggtggccaatatcatcttggctcgatcctggcaccaaaatcaatc	364					
QY	361	aatcggagca	ttcatacgtgtcgggactcgtgtcgttcttcatggatcgctatcgga	420					
Db	365	aatcggagca	ttcatacgtgtcgggactcgtgtcgttcttcatggatcgctatcgga	424					
QY	421	agctcagtg	accatgatacctgctcaacagatcccgcgctgatagccgcctatagtcac	480					
Db	425	agctcagtg	accatgatacctgctcaacagatcccgcgctgatagccgcctatagtcac	484					
QY	481	acatactcc	aggattcgtggcattgtcttccctcatctacgaggagtcgcgtggtg	540					
Db	485	acatactcc	aggattcgtggcattgtcttccctcatctacgaggagtcgcgtggtg	544					
QY	541	ctctgttca	actgttttgcctgaaagaaaccccttgggttcattgaaatcggattcggc	600					
Db	545	ctctgttca	actgttttgcctgaaagaaaccccttgggttcattgaaatcggattcggc	604					
QY	601	ctcacgata	ctctctctcgtgatctccttattcgcacattctagcgcggtccaaacccaag	660					
Db	605	ctcacgata	ctctctctcgtgatctccttattcgcacattctagcgcggtccaaacccaag	664					
QY	661	gtccaaaac	tcaacaggtatggactgcttggagcaactatactggctggtccgacgctc	720					
Db	665	gtccaaaac	tcaacaggtatggactgcttggagcaactatactggctggtccgacgctc	724					
QY	721	tgcttcata	ctctggccttctcgacatcctgcttcatgttcaattggcttggacgcagcaatg	780					
Db	725	tgcttcata	ctctggccttctcgacatcctgcttcatgttcaattggcttggacgcagcaatg	784					
QY	781	catctggct	gaagaatgcacagatgctgctcgtacgggtacccaaagcagtggtcagtgca	840					
Db	785	catctggct	gaagaatgcacagatgctgctcgtacgggtacccaaagcagtggtcagtgca	844					
QY	841	atcataa	attggtctctgcaccgccttcccatatacaatcgcagttctgtatggaaattaca	900					
Db	845	atcataa	attggtctctgcaccgccttcccatatacaatcgcagttctgtatggaaattaca	904					
QY	901	gatctcgac	tctattctaaagtcccgcgcttatattccattcgcagacaatgcacgcagtc	960					
Db	905	gatctcgac	tctattctaaagtcccgcgcttatattccattcgcagacaatgcacgcagtc	964					
QY	961	cttcgggtc	gtcagttttgtcaacggtcctctcattgtggcggtatcgtgatggcctcttc	1020					
Db	965	cttcgggtc	gtcagttttgtcaacggtcctctcattgtggcggtatcgtgatggcctcttc	1024					
QY	1021	gccctcaac	gcgtgtacaagagactgctctcgactcaactggagcttttgcaccgggacaat	1080					

Db	1025	gcccctcaac	gcgtgtacaagagactgcgtctcgactcacctggagctttgcccgggacaat	1084					
QY	1081	gggctgggtat	tttccactcatctcgaaacgcatcattcatccccgcgtggcaagtctcctgttg	1140					
Db	1085	gggctgggtat	tttccactcatctcgaaacgcatcattcatccccgcgtggcaagtctcctgttg	1144					
QY	1141	tctctattc	gcgcacctggggaattctggccacatgcggatgtatatattcttaggttctagc	1200					
Db	1145	tctctattc	gcgcacctggggaattctggccacatgcggatgtatatattcttaggttctagc	1204					
QY	1201	acagctttc	aatgccttgggtcaattccgcggttggtactccagcaactctccttctctgac	1260					
Db	1205	acagctttc	aatgccttgggtcaattccgcggttggtactccagcaactctccttctctgac	1264					
QY	1261	ccaatgc	ccctactcctctacccaaaagcgagatccaaaagtcttgcgcgagcactcgtgct	1320					
Db	1265	ccaatgc	ccctactcctctacccaaaagcgagatccaaaagtcttgcgcgagcactcgtgct	1324					
QY	1321	tttgtgtt	accgcgtgggaatcggtttctgggtcaatgtgctagcgggtggtcttcacgtcc	1380					
Db	1325	tttgtgtt	accgcgtgggaatcggtttctgggtcaatgtgctagcgggtggtcttcacgtcc	1384					
QY	1381	gtcaccac	tgtgttttcaagcttccactgacogtgcctacggcgcgtcaaccatgaat	1440					
Db	1385	gtcaccac	tgtgttttcaagcttccactgacogtgcctacggcgcgtcaaccatgaat	1444					
QY	1441	tacacaag	tgcgattatagcgttgcacttgcctcttgggtccttgaactgggtcgtgcat	1500					
Db	1445	tacacaag	tgcgattatagcgttgcacttgcctcttgggtccttgaactgggtcgtgcat	1504					
QY	1501	gccaggaa	gcattatcaggggaccccaacttggagccttgacggagcgggtcgtcggagcagaa	1560					
Db	1505	gccaggaa	gcattatcaggggaccccaacttggagccttgacggagcgggtcgtcggagcagaa	1564					
QY	1561	tttcaagt	ttgggccatga	1578					
Db	1565	tttcaagt	ttgggccatga	1582					

RESULT 13
US-09-351-224-7
; Sequence 7, Application US/09351224
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jon
; TITLE OF INVENTION: Compositions and Methods for Fummonisin Detoxification
; FILE REFERENCE: 5718-111
; CURRENT APPLICATION NUMBER: US/09/351.224
; CURRENT FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 7
; LENGTH: 1578
; TYPE: DNA
; ORGANISM: Exophiala spinifera
; FEATURE:
; OTHER INFORMATION: permease, coding sequence
US-09-351-224-7

Query Match 98.6%; Score 1555.6; DB 17; Length 1578; Best Local Similarity 99.1%; Pred. No. 0; Matches 1564; Conservative 0; Mismatches 14; Indels 0; Gaps 0;									
QY	1	atggactcc	agaccacaaagtggatacggcgagaaaaggcggaacaaggcagacaacgaagaac	60					
Db	1	atggactcc	agaccacaaagtggatacggcggttgaaggcgggacaaggcagacaacgaagaac	60					
QY	61	acagagac	ggcgcgaggtggtgcgtccagagtccttgaacgttctctctggagaagaaa	120					
Db	61	acagagac	ggcgcgaggtggtgcgtccagagtccttgaacgttctctctggagaagaaa	120					
QY	121	caatttgg	caccatcaccatcgtctccttggcctttgtgatttgcaacagttgggctggt	180					

Db 301 gtctaccgactgccggtggccaaatatcatcttttgcgtcgatccctggcaccacaaatacaatc 360
QY 361 aatcggagcatttcatacagtggtcgggactcgtgtcggttgctttcattcgatcgcgtatcgga 420
Db 361 aatcggagcatttcatacagtggtcgggactcgtgtcggttgctttcattcgatcgcgtatcgga 420
QY 421 agctcagtgaccatgataccctgctcaacagatcccgcgctgatagcgcgcctatagtca 480
Db 421 agctcagtgaccatgataccctgctcaacagatcccgcgctgatagcgcgcctatagtca 480
QY 481 acatactccaggattcgtggcatgtcttctctcatctacgaggagtcgcgtggtg 540
Db 481 acatactccaggattcgtggcatgtcttctctcatctacgaggagtcgcgtggtg 540
QY 541 ctcttgttcaactgtttgtccctgaaagaaaccccttgggttcattgaaatcggttcggc 600
Db 541 ctcttgttcaactgtttgtccctgaaagaaaccccttgggttcattgaaatcggttcggc 600
QY 601 ctacagatcgtctctctcgtgatctctctttatgccattctagcgcgggtccaccccaag 660
Db 601 ctacagatcgtctctctcgtgatctctctttatgccattctagcgcgggtccaccccaag 660
QY 661 gctccaaactcaacagggtatggactgcttgagagcaactatactggtggtccgcagcgctc 720
Db 661 gctccaaactcaacagggtatggactgcttgagagcaactatactggtggtccgcagcgctc 720
QY 721 tgccttcactcgtggccctttgcacatcctgcttcattggttcattggtggacgcagcaatg 780
Db 721 tgccttcactcgtggccctttgcacatcctgcttcattggttcattggtggacgcagcaatg 780
QY 781 catctggctgaagaatgcacagatgctgctcgtacggtacccaaagcagtggttcagtga 840
Db 781 catctggctgaagaatgcacagatgctgctcgtacggtacccaaagcagtggttcagtga 840
QY 841 atcataattggcttctgcaccgcctttcccatatacaaatcgacgttctgtatggaattaca 900
Db 841 atcataattggcttctgcaccgcctttcccatatacaaatcgacgttctgtatggaattaca 900
QY 901 gatctcgactctattctaaagtctccgcttatattccattcgagacaatgacgcagtc 960
Db 901 gatctcgactctattctaaagtctccgcttatattccattcgagacaatgacgcagtc 960
QY 961 ctctcggtcgtcagttttgcaacgggtcctctcattggtgggtatcgtgatgccttcttc 1020
Db 961 gctcgaaattcggctttttgcaacgggtcctctcattggtgggtatcgtgatgccttcttc 1020
QY 1021 gccctcaacgctgtacaaagagactcgtctcgcactcaactggagctttgcccgggacaaat 1080
Db 1021 gccctcaacgctgtacaaagagactcgtctcgcactcaactggagctttgcccgggacaaat 1080
QY 1081 gggctggtattttccactcatctcgaaacgcattatccccgcgtggcaagttcctgtttgg 1140
Db 1081 gggctggtattttccactcatctcgaaacgcattatccccgcgtggcaagttcctgtttgg 1140
QY 1141 tctctattcgcgacctggggaaattctggccacatgcgggatgtatatattctaggttttagc 1200
Db 1141 tctctattcgcgacctggggaaattctggccacatgcgggatgtatatattctaggttttagc 1200
QY 1201 acagctttcaatgccttgggtcaattccgcggttgcactccagcaactctcttccctgac 1260
Db 1201 acagctttcaatgccttgggtcaattccgcggttgcactccagcaactctcttccctgac 1260
QY 1261 ccaatcgccctactcctctacacaaagcgagatccaaaagtcttgcgagcactcgtgct 1320
Db 1261 ccaatcgccctactcctctacacaaagcgagatccaaaagtcttgcgagcactcgtgct 1320
QY 1321 ttgtgttacgcgctggaaatcggtttcttggtcaatgtgctagcgggtggtcttcacgtcc 1380
Db 1321 ttgtgttacgcgctggaaatcggtttcttggtcaatgtgctagcgggtggtcttcacgtcc 1380
QY 1381 gtcaccactgtgttttccagcttcccactgacccgtgcctacggtgcgggtcaaccatgaat 1440
Db 1381 gtcaccactgtgttttccagcttcccactgacccgtgcctacggtgcgggtcaaccatgaat 1440

QY 1441 tacacaagtgcgattatagcggttgcaacttgctcttggtgtcttgaactgggtcgtgcat 1500
Db 1441 tacacaagtgcgattatagcggttgcaacttgctcttggtgtcttgaactgggtcgtgcat 1500
QY 1501 gccaggagcattatcagggagcccccacttggagcttgacggagcggtcgtcgagcagaa 1560
Db 1501 gccaggagcattatcagggagcccccacttggagcttgacggagcggtcgtcgagcagaa 1560
QY 1561 tttcaagtgggccaatga 1578
Db 1561 tttcaagtgggccaatga 1578

RESULT 15
US-09-677-488-7
; Sequence 7, Application US/09677488
; GENERAL INFORMATION:
; APPLICANT: DuVick, Jon
; TITLE OF INVENTION: Compositions and Methods for Fummonisin Detoxification
; FILE REFERENCE: 35718/204100 (5718-111A)
; CURRENT APPLICATION NUMBER: US/09/677,488
; CURRENT FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 1578
; TYPE: DNA
; ORGANISM: Exophiala spinifera
; FEATURE:
; OTHER INFORMATION: permease, fully spliced cDNA
US-09-677-488-7

Query Match 98.6%; Score 1555.6; DB 26; Length 1578;
Best Local Similarity 99.1%; Pred. NO. 0;
Matches 1564; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 1 atggactccagaccaaagtggatagcggtgagaaaggcgagacaaaggcagacaaacgaagaac 60
Db 1 atggactccagaccaaagtggatagcggtgagaaaggcgagacaaaggcagacaaacgaagaac 60
QY 61 acagagacggcgggcgaggttggtgcgtccgagtcacctgaacgttccctcgtggagaagaaa 120
Db 61 acagagagggcgggcgaggttggtgcgtccgagtcacctgaacgttccctcgtggagaagaaa 120
QY 121 caatttggcaccatcaccaatcgtgtccttggccttggatttgcaacagttgggctggt 180
Db 121 caatttggcaccatcaccaatcgtgtccttggccttggatttgcaacagttgggctggt 180
QY 181 atctcaggcagtgctccagctcgcacctactagcggggggcccgctcactctcctttacggc 240
Db 181 atctcaggcagtgctccagctcgcacctactagcggggggcccgctcactctcctttacggc 240
QY 241 atcctaatacagtgatctcgtctacatctgcacatcgtcttctcattagcgaactgaccagc 300
Db 241 atcctaatacagtgatctcgtctacatctgcacatcgtcttctcattagcgaactgaccagc 300
QY 301 gtctacccgactgccgggtggccaaatatcatcttggcgtcgatccctggcaccacaaatacaatc 360
Db 301 gtctacccgactgccgggtggccaaatatcatcttggcgtcgatccctggcaccacaaatacaatc 360
QY 361 aatcggagcatttcatacagtggtgggactcgtgctgttctcatggatcgctatcgga 420
Db 361 aatcggagcatttcatacagtggtgggactcgtgctgttctcatggatcgctatcgga 420
QY 421 agctcagtgaccatgataccctgctcaacagatcccgcgctgatagccgcctatagtcac 480
Db 421 agctcagtgaccatgataccctgctcaacagatcccgcgctgatagccgcctatagtcac 480
QY 481 acatactccaggattcgtggcatgtcttctctcatctacgaggagtcgcgtggtggtg 540

```

Db      1561 tttcaagttgggccaatga 1578

Search completed: April 27, 2002, 01:34:27
Job time: 22031 sec

|||||
481 acatactccacaggattcgtggcatgtcttctcatctacagggagtcgcgtggtg 540
QY      541 ctcttggttcaactgttttgccttgaaagaaaccccttgggttcatgaaatcggattcggc 600
Db      541 ctcttggttcaactgttttgccttgaaagaaaccccttgggttcatgaaatcggattcggc 600
QY      601 ctacgatacgtctctcttctcgtgatctctcttctacgccaattctagcgcggctccaaccccaag 660
Db      601 ctacgatacgtctctcttctcgtgatctctcttctacgccaattctagcgcggctccaaccccaag 660
QY      661 gctccaaaactcacagggtatggactgcttggagcaactatatacttggctggtccgacggcgtc 720
Db      661 gctccaaaactcacagggtatggactgcttggagcaactatatacttggctggtccgacggcgtc 720
QY      721 tgcctcatctctgggccccttctcgacatccctgcttcatgttcatgttggacgcagcaatg 780
Db      721 tgcctcatctctgggccccttctcgacatccctgcttcatgttcatgttggacgcagcaatg 780
QY      781 catctggctgaagaatgcacagatgctcgtctacggtaacccaaagcagtggtcagtgca 840
Db      781 catctggctgaagaatgcacagatgctcgtctacggtaacccaaagcagtggtcagtgca 840
QY      841 atcataattggcttctgcacccgcttccatatacataatgcagcttctgtatggaattaca 900
Db      841 atcataattggcttctgcacccgcttccatatacataatgcagcttctgtatggaattaca 900
QY      901 gatctcgactctatttctaagttccgcggcttatccatccagacaaatgcagcagtc 960
Db      901 gatctcgactctatttctaagttccgcggcttatccatccagacaaatgcagcagtc 960
QY      961 ctctcgtcgtcagtttttgcaacgggtccctctcctcgtggtggtatcgtgacgcttcttc 1020
Db      961 gctcgaaattcggctttttgcaacgggtccctctcctcgtggtggtatcgtgacgcttcttc 1020
QY      1021 gccctcaacgctgtacaagagactgctctcgactcaactggagctttgcccggggacaat 1080
Db      1021 gccctcaacgctgtacaagagactgctctcgactcaactggagctttgcccggggacaat 1080
QY      1081 gggctgggtattttccactcatctcgaacgcattcattcccccgtggcaagtctctgtttgg 1140
Db      1081 gggctgggtattttccactcatctcgaacgcattcattcccccgtggcaagtctctgtttgg 1140
QY      1141 tctctattcgcgacctgggggaattcttggccacatcgcgatgtatatctctaggttcttagc 1200
Db      1141 tctctattcgcgacctgggggaattcttggccacatcgcgatgtatatctctaggttcttagc 1200
QY      1201 acagctttcaatgccttggtaattccgcctgtgtactccagcaactcctcctcctgac 1260
Db      1201 acagctttcaatgccttggtaattccgcctgtgtactccagcaactcctcctcctgac 1260
QY      1261 ccaatcgccctactcctctacaaaagcgagatccaaagtcttgcgaggaactcgtgct 1320
Db      1261 ccaatcgccctactcctctacaaaagcgagatccaaagtcttgcgaggaactcgtgct 1320
QY      1321 ttgtgttaccgcgtggaaatcggtttctggtcaatgtgctagcgggtggtcttcaacgtcc 1380
Db      1321 ttgtgttaccgcgtggaaatcggtttctggtcaatgtgctagcgggtggtcttcaacgtcc 1380
QY      1381 gtcaccactgtgtttttcagcttcccactgacctgacctacggccgctcaacccatgaat 1440
Db      1381 gtcaccactgtgtttttcagcttcccactgacctgacctacggccgctcaacccatgaat 1440
QY      1441 tacacaaatgcgattatagggcgttgcacttgctcttggtgtcttgaaactgggtcgtgcat 1500
Db      1441 tacacaaatgcgattatagggcgttgcacttgctcttggtgtcttgaaactgggtcgtgcat 1500
QY      1501 gccagggaagcattatcaggggaaccccaacttggagcttgacggagcgggtcgtcggagcagaa 1560
Db      1501 gccagggaagcattatcaggggaaccccaacttggagcttgacggagcgggtcgtcggagcagaa 1560
QY      1561 tttcaagttgggccaatga 1578
|||||

```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2002, 01:35:42 ; Search time 150.7 Seconds
(without alignments)
2572.062 Million cell updates/sec

Title: US-09-882-694A-7
Perfect score: 1578
Sequence: 1 atggactccagaccgaagtgg.....aatttcaagttgggcatga 1578

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	58	3.7	7218	1	US-08-232-463-14
2	35.8	2.3	289	4	US-09-007-005-17
3	35.8	2.3	289	4	US-09-244-796-17
4	34	2.2	1473	2	US-08-541-033A-25
5	34	2.2	1473	2	US-08-828-451-25
6	34	2.2	1506	2	US-08-541-033A-23
7	34	2.2	1506	2	US-08-828-451-23
8	34	2.2	1969	2	US-08-541-033A-7
9	34	2.2	1969	2	US-08-828-451-7
10	34	2.2	2096	2	US-08-541-033A-19
11	34	2.2	2096	2	US-08-828-451-19
12	34	2.2	2099	2	US-08-541-033A-3
13	34	2.2	2099	2	US-08-828-451-3
14	34	2.2	2137	2	US-08-541-033A-18
15	34	2.2	2137	2	US-08-828-451-18
16	34	2.2	2140	2	US-08-541-033A-1
17	34	2.2	2140	2	US-08-828-451-1
18	32.8	2.1	642	4	US-09-199-637A-26
19	32.8	2.1	957	4	US-09-199-637A-22
20	32.8	2.1	1407	4	US-09-199-637A-28
21	32.8	2.1	3141	4	US-09-199-637A-12
22	32.8	2.1	42235	4	US-09-199-637A-1
23	32.6	2.1	686	4	US-09-328-111-364
24	32.2	2.0	1249	3	US-09-113-536-1
25	32.2	2.0	1249	5	PCT-US95-05785-1
26	31.6	2.0	2102	1	US-08-784-651-5
27	31.2	2.0	44377	2	US-08-804-227C-7

c 28	31.2	2.0	44377	2	US-08-804-198-1	Sequence 1, Appli
c 29	31	2.0	192	1	US-08-030-731A-45	Sequence 45, Appl
c 30	31	2.0	192	1	US-08-030-731A-46	Sequence 46, Appl
	31	1.9	1279	1	US-08-146-010A-4	Sequence 4, Appli
c 32	30.4	1.9	1279	1	US-08-674-168-9	Sequence 9, Appli
c 33	30.4	1.9	12047	2	US-09-022-461-1	Sequence 1, Appli
c 34	30	1.9	702	4	US-08-843-417-3	Sequence 3, Appli
c 35	30	1.9	947	1	US-08-474-177-36	Sequence 36, Appl
c 36	30	1.9	947	1	US-08-487-033-36	Sequence 36, Appl
c 37	30	1.9	947	1	US-08-480-810-36	Sequence 36, Appl
c 38	30	1.9	947	2	US-08-508-735-36	Sequence 36, Appl
c 39	30	1.9	947	2	US-08-848-251-36	Sequence 36, Appl
c 40	30	1.9	947	2	US-08-486-047-36	Sequence 36, Appl
c 41	30	1.9	947	3	US-09-120-130-36	Sequence 36, Appl
c 42	30	1.9	947	3	US-09-115-252-36	Sequence 36, Appl
c 43	30	1.9	947	3	US-08-986-515-36	Sequence 36, Appl
c 44	30	1.9	947	3	US-09-120-128-36	Sequence 36, Appl
c 45	30	1.9	947	4	US-09-120-129-36	Sequence 36, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: PT29pt-Fls
US-08-232-463-14

;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/828,451
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/541,033
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF155
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2096 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; US-08-828-451-19

Query Match 2.2%; Score 34; DB 2; Length 2096;
Best Local Similarity 56.1%; Pred. No. 0.31;
Matches 64; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

Qy 1460 gcgttgcaattgctctgtgtgtcttgaactgggtcgtgcatgccaggaagcattatcagg 1519
||||| ||||||| ||||||| ||| ||| ||| ||| ||| |||
Db 640 GCCTCGCTCTTGCCCTTGGGTCGAAGTCGGAGCGCCCTTGCCGCGGCCCATGGGCAGG 581

Qy 1520 gaccccaacttgagcttgacggacgggtcgtcgtagcagaaatttcaagttgggc 1573
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 580 GTGGTCAGGCTGTTCTTGAAGATCTGCTCAAGGCAAGGAACCTTCATGATGGAC 527

RESULT 12
US-08-541-033A-3/c
; Sequence 3, Application US/08541033A
; Patent No. 5879941
; GENERAL INFORMATION:
; APPLICANT: Schmidt, Robert R.
; APPLICANT: Miller, Philip
; TITLE OF INVENTION: NOVEL POLYPEPTIDES AND POLYNUCLEOTIDES
; TITLE OF INVENTION: RELATING TO THE '-' AND -SUBUNITS OF GLUTAMATE
; TITLE OF INVENTION: DEHYDROGENASES AND METHODS OF USE
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: Florida
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/541,033A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF155
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:

;
; LENGTH: 2099 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 33..1568
; US-08-541-033A-3

Query Match 2.2%; Score 34; DB 2; Length 2099;
Best Local Similarity 56.1%; Pred. No. 0.31;
Matches 64; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

Qy 1460 gcgttgcaattgctctgtgtgtcttgaactgggtcgtgcatgccaggaagcattatcagg 1519
||||| ||||||| ||||||| ||| ||| ||| ||| ||| |||
Db 643 GCCTCGCTCTTGCCCTTGGGTCGAAGTCGGAGCGCCCTTGCCGCGGCCCATGGGCAGG 584

Qy 1520 gaccccaacttgagcttgacggacgggtcgtcgtagcagaaatttcaagttgggc 1573
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 583 GTGGTCAGGCTGTTCTTGAAGATCTGCTCAAGGCAAGGAACCTTCATGATGGAC 530

RESULT 13
US-08-828-451-3/c
; Sequence 3, Application US/08828451
; Patent No. 5985634
; GENERAL INFORMATION:
; APPLICANT: Schmidt, Robert R.
; APPLICANT: Miller, Philip
; TITLE OF INVENTION: NOVEL POLYPEPTIDES AND POLYNUCLEOTIDES
; TITLE OF INVENTION: RELATING TO THE '-' AND -SUBUNITS OF GLUTAMATE
; TITLE OF INVENTION: DEHYDROGENASES AND METHODS OF USE
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: Florida
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/828,451
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/541,033
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF155
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2099 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 33..1568
; US-08-828-451-3

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2002, 00:40:59 ; Search time 706.01 Seconds
(without alignments)
3837.468 Million cell updates/sec

Title: us-09-882-694A-7
Perfect score: 1578
Sequence: 1 atggactccagaccaaagtg.....aatttcaagttggccatga 1578

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802.*
1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1555.6	98.6	1764	22	AAD02693
2	46.4	2.9	724	21	AAF12818
3	40.6	2.6	552	21	AAF12587
4	39.2	2.5	1551	21	AAC49980
5	39.2	2.5	1554	21	AAC43230
6	39.2	2.5	1664	21	AAC49990
7	39.2	2.5	5839	19	AAV46335
8	38.2	2.4	6741	21	AA10595
9	37.8	2.4	3510	16	AAQ92829
					Exophiala spinifera
					Aspergillus oryzae
					Aspergillus oryzae
					Arabidopsis thalia
					Arabidopsis thalia
					Arabidopsis thalia
					Hydra head activat
					Gene encoding a su
					Mitochondrial cyto

10	37.6	2.4	1686	21	AAC38835	Arabidopsis thalia
11	36.6	2.3	10732	21	AA10594	Gene encoding a su
12	36.6	2.3	99960	21	AAZ50905	Human TBC-1 partia
13	36.2	2.3	1900	12	AAQ14834	Cytochrome P450C25
14	35.4	2.2	2036	21	AAF07981	Fusarium venenatum
15	35	2.2	4590	22	AAH24065	Yeast AOD9604-asso
16	34.6	2.2	446	22	AA114748	Human breast cance
17	34.6	2.2	550	22	AA123610	Human breast cance
18	34.4	2.2	291	21	AAZ90043	Hydrophobic domain
19	34.4	2.2	341	21	AAZ42802	Human 5' EST isola
20	34.4	2.2	677	22	AAH07357	Human cDNA clone (
21	34.4	2.2	1045	22	AAH14780	Human cDNA sequenc
22	34.4	2.2	1189	21	AAZ90053	Hydrophobic domain
23	34.4	2.2	5059	20	AAZ84332	Stealth virus nucl
24	34.2	2.2	15359	22	ABA08724	Human ryanodine re
25	34.2	2.2	15359	23	AAS69156	DNA encoding novel
26	34	2.2	627	22	AAI98766	Human excretory re
27	34	2.2	627	22	AA163162	Human kidney relat
28	34	2.2	1473	18	AA164548	NADP-specific glut
29	34	2.2	1506	18	AA164547	NADP-specific glut
30	34	2.2	1969	18	AA164531	Glutamate dehydrog
31	34	2.2	2096	18	AA164543	NADP-specific glut
32	34	2.2	2099	18	AA164530	NADP-specific glut
33	34	2.2	2137	18	AA164542	NADP-specific glut
34	34	2.2	2140	18	AA164529	NADP-specific glut
35	34	2.2	5331	21	AAA30292	DNA fragment carry
36	33.8	2.1	3367	23	ABL02489	Drosophila melanog
37	33.8	2.1	10997	23	ABL02488	Drosophila melanog
38	33.6	2.1	2724	22	ABA15116	Human nervous syst
39	33.2	2.1	410	22	AAS36979	Novel human diagno
40	33.2	2.1	2127	23	AAS88963	DNA encoding novel
41	33.2	2.1	3168	23	ABL12363	Drosophila melanog
42	33	2.1	1038	23	AAS73881	DNA encoding novel
43	33	2.1	1109	23	ABL03311	Drosophila melanog
44	33	2.1	1977	22	AAK53275	Human polynucleoti
45	33	2.1	1978	23	AAS73883	DNA encoding novel

ALIGNMENTS

RESULT 1

AAD02693

ID AAD02693 standard; DNA; 1764 BP.

XX

AC AAD02693;

XX 02-MAY-2001 (first entry)

DT

XX Exophiala spinifera permease DNA.

DE

XX Fumonisin; degradative enzyme; transport enzyme; flavin monooxygenase;

KW aldehyde dehydrogenase; permease; p-glycoprotein; fumonisin esterase;

KW detoxification; mycotoxin; animal feed; human feed; silage;

KW transgenic plant; transgenic animal; microbial spray; ds.

XX Exophiala spinifera.

OS

XX Key

FH CDS

FT Location/Qualifiers

FT 5..1582

FT /tag= a

FT /product= "E. spinifera permease"

FT /note= "This region is specifically claimed in

claim 1b as SEQ ID NO: 7"

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

QY 1561 ttccaagtgggccaatga 1578
|||||
Db 1565 ttccaagtgggccaatga 1582

RESULT 2
AAFL12818
ID AAF12818 standard; CDNA; 724 BP.
XX
AC AAF12818;
XX
DT 13-MAR-2001 (first entry)
XX
DE Aspergillus oryzae EST SEQ ID NO:5341.
XX

KW Multiple gene expression; filamentous fungal cell; EST;
KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW culture condition; environmental stress; spore morphogenesis;
KW metabolic pathway engineering; catabolic pathway engineering; ss.
XX
OS Aspergillus oryzae.
XX
PN WO200056762-A2.
XX
PD 28-SEP-2000.
XX
PF 22-MAR-2000; 2000WO-US07781.
XX
PR 22-MAR-1999; 99US-0273623.
XX

PA (NOVO) NOVO NORDISK BIOTECH INC.
PA (NOVO) NOVO NORDISK AS.
XX
PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX
DR WPI; 2000-594572/56.
XX

PT Monitoring differential expression of genes in filamentous fungal cells
PT uses fluorescence-labeled nucleic acids isolated from the cells and a
PT substrate of expressed sequence tags -
XX
PS Claim 88; Page 2226; 3161pp; English.
XX

CC The present invention describes a method for monitoring differential
CC expression of genes in a first filamentous fungal (FF) cell relative to
CC expression of the same genes in one or more second filamentous fungal
CC cells. The method uses fluorescence-labeled nucleic acids isolated from
CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC are used in the methods for monitoring differential expression of genes
CC in a first filamentous fungal (FF) cell relative to expression of the
CC same genes in one or more second filamentous fungal cells. Monitoring
CC the global expression of genes from FF cells allows the production
CC potential of the microorganisms to be improved. New genes may be
CC discovered, possible functions of unknown open reading frames can be
CC identified and gene copy number variation and stability can be
CC monitored. The expression of genes can be used to study how FF cells
CC adapt to changes in culture conditions, environmental stress, spore
CC morphogenesis, recombination, metabolic or catabolic pathway
CC engineering. Using ESTs provides several advantages over genomic or
CC random cDNA clones including elimination of redundancy as one spot on an
CC array equals one gene or open reading frame, and organisation of the
CC microarrays based on function of the gene products to facilitate
CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
CC all specifically claimed in the present invention.
XX
SQ Sequence 724 BP; 153 A; 221 C; 175 G; 175 T; 0 other;

Query Match 2.9%; Score 46.4; DB 21; Length 724;
Best Local Similarity 46.7%; Pred. No. 0.00031;
Matches 183; Conservative 0; Mismatches 206; Indels 3; Gaps 1;

QY 757 ttcatgggttgagcagcaatgcatctggctgaagaatgcacagatgctgctgacg 816
|||
Db 332 ttagctggcctcgacggagtaatccacatggcgaagaagtcaaaaaagcgaagtgtgc 391
|||
QY 817 gtaccaaaagcagtggtcagtgcaatcataatggtcttgcacccgctttccatataca 876
|||
Db 392 gtgcacggggccatgatctctgcaccatgatcaacggaaccccttgcttgcctacctc 451
|||
QY 877 atgcagttctgtatggaattacagatct--cgactctattcttaagtccggcttat 933
|||||
Db 452 atgcagacctgtactgcattggcgactacacccgaggcgtgctgagcccccagggtac 511
|||
QY 934 attcattcgagacaatgacgagtccttcccttgcgtcagtttgcaacggtcctctca 993
|||
Db 512 cgcatcatcacaatgcctaccagccacgggtctccaaagcagcaacctacgtgatg 571
|||
QY 994 tgtggcggtatcgatggccttcttcgcccctcaacgctgtacaagagactgcttcga 1053
|||
Db 572 gcgatgggcatgctcccggtggtgctgcccctctttaacggtctgcctccgtaaccgc 631
|||
QY 1054 ctcaectggagctttgcccgggacaatgggctggttatcttccactcatctcgaaacgcaatt 1113
|||||
Db 632 ctcaectgggcatctgcacgcgataacggcctccctcttctgacttcttcgcccgcgtc 691
|||
QY 1114 catcccgctggcaagtctctgtttggtctct 1145
|||
Db 692 gacccaaacctataataatccccctcgcgctct 723
|||

RESULT 3
AAFL12587
XX ID AAF12587 standard; CDNA; 552 BP.
XX
AC AAF12587;
XX
DT 13-MAR-2001 (first entry)
XX
DE Aspergillus oryzae EST SEQ ID NO:5110.
XX
KW Multiple gene expression; filamentous fungal cell; EST;
KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW culture condition; environmental stress; spore morphogenesis;
KW metabolic pathway engineering; catabolic pathway engineering; ss.
XX
OS Aspergillus oryzae.
XX
PN WO200056762-A2.
XX
PD 28-SEP-2000.
XX
PF 22-MAR-2000; 2000WO-US07781.
XX
PR 22-MAR-1999; 99US-0273623.
XX
PA (NOVO) NOVO NORDISK BIOTECH INC.
PA (NOVO) NOVO NORDISK AS.
XX
PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX
DR WPI; 2000-594572/56.
XX

PT Monitoring differential expression of genes in filamentous fungal cells
PT uses fluorescence-labeled nucleic acids isolated from the cells and a
PT substrate of expressed sequence tags -
XX
PS Claim 88; Page 2142; 3161pp; English.
XX
CC The present invention describes a method for monitoring differential

PF	25-FEB-2000;	2000EP-0301439.			
XX					
PR	25-FEB-1999;	99US-0121825.	PR	09-JUL-1999;	99US-0142920.
PR	05-MAR-1999;	99US-0123180.	PR	12-JUL-1999;	99US-0142977.
PR	09-MAR-1999;	99US-0123548.	PR	13-JUL-1999;	99US-0143542.
PR	23-MAR-1999;	99US-0125788.	PR	14-JUL-1999;	99US-0143624.
PR	25-MAR-1999;	99US-0126264.	PR	15-JUL-1999;	99US-0144005.
PR	29-MAR-1999;	99US-0126785.	PR	16-JUL-1999;	99US-0144085.
PR	01-APR-1999;	99US-0127462.	PR	16-JUL-1999;	99US-0144086.
PR	06-APR-1999;	99US-0128234.	PR	19-JUL-1999;	99US-0144325.
PR	08-APR-1999;	99US-0128714.	PR	19-JUL-1999;	99US-0144331.
PR	16-APR-1999;	99US-0129845.	PR	19-JUL-1999;	99US-0144332.
PR	19-APR-1999;	99US-0130077.	PR	19-JUL-1999;	99US-0144333.
PR	21-APR-1999;	99US-0130449.	PR	19-JUL-1999;	99US-0144334.
PR	23-APR-1999;	99US-0130510.	PR	19-JUL-1999;	99US-0144335.
PR	28-APR-1999;	99US-0130891.	PR	20-JUL-1999;	99US-0144352.
PR	30-APR-1999;	99US-0131449.	PR	20-JUL-1999;	99US-0144632.
PR	30-APR-1999;	99US-0132048.	PR	21-JUL-1999;	99US-0144884.
PR	04-MAY-1999;	99US-0132484.	PR	21-JUL-1999;	99US-0144814.
PR	05-MAY-1999;	99US-0132407.	PR	21-JUL-1999;	99US-0145086.
PR	06-MAY-1999;	99US-0132485.	PR	21-JUL-1999;	99US-0145088.
PR	06-MAY-1999;	99US-0132486.	PR	22-JUL-1999;	99US-0145085.
PR	07-MAY-1999;	99US-0132863.	PR	22-JUL-1999;	99US-0145087.
PR	11-MAY-1999;	99US-0134256.	PR	22-JUL-1999;	99US-0145089.
PR	14-MAY-1999;	99US-0134218.	PR	22-JUL-1999;	99US-0145192.
PR	14-MAY-1999;	99US-0134219.	PR	23-JUL-1999;	99US-0145145.
PR	14-MAY-1999;	99US-0134221.	PR	23-JUL-1999;	99US-0145218.
PR	14-MAY-1999;	99US-0134370.	PR	23-JUL-1999;	99US-0145224.
PR	18-MAY-1999;	99US-0134768.	PR	26-JUL-1999;	99US-0145276.
PR	19-MAY-1999;	99US-0134941.	PR	27-JUL-1999;	99US-0145913.
PR	20-MAY-1999;	99US-0135124.	PR	27-JUL-1999;	99US-0145918.
PR	21-MAY-1999;	99US-0135353.	PR	27-JUL-1999;	99US-0145919.
PR	24-MAY-1999;	99US-0135629.	PR	28-JUL-1999;	99US-0145951.
PR	25-MAY-1999;	99US-0136021.	PR	02-AUG-1999;	99US-0146386.
PR	27-MAY-1999;	99US-0136392.	PR	02-AUG-1999;	99US-0146388.
PR	28-MAY-1999;	99US-0136782.	PR	02-AUG-1999;	99US-0146389.
PR	01-JUN-1999;	99US-0137222.	PR	03-AUG-1999;	99US-0147038.
PR	03-JUN-1999;	99US-0137528.	PR	04-AUG-1999;	99US-0147204.
PR	04-JUN-1999;	99US-0137502.	PR	04-AUG-1999;	99US-0147302.
PR	07-JUN-1999;	99US-0137724.	PR	05-AUG-1999;	99US-0147192.
PR	08-JUN-1999;	99US-0138094.	PR	06-AUG-1999;	99US-0147260.
PR	10-JUN-1999;	99US-0138540.	PR	06-AUG-1999;	99US-0147303.
PR	10-JUN-1999;	99US-0138847.	PR	06-AUG-1999;	99US-0147416.
PR	14-JUN-1999;	99US-0139119.	PR	09-AUG-1999;	99US-0147493.
PR	16-JUN-1999;	99US-0139452.	PR	09-AUG-1999;	99US-0147935.
PR	16-JUN-1999;	99US-0139453.	PR	10-AUG-1999;	99US-0148171.
PR	17-JUN-1999;	99US-0139492.	PR	11-AUG-1999;	99US-0148319.
PR	18-JUN-1999;	99US-0139454.	PR	12-AUG-1999;	99US-0148341.
PR	18-JUN-1999;	99US-0139455.	PR	13-AUG-1999;	99US-0148565.
PR	18-JUN-1999;	99US-0139456.	PR	13-AUG-1999;	99US-0148684.
PR	18-JUN-1999;	99US-0139457.	PR	16-AUG-1999;	99US-0149368.
PR	18-JUN-1999;	99US-0139458.	PR	17-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	99US-0139459.	PR	18-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	99US-0139460.	PR	20-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	99US-0139461.	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	99US-0139462.	PR	20-AUG-1999;	99US-0149929.
PR	18-JUN-1999;	99US-0139463.	PR	23-AUG-1999;	99US-0149902.
PR	18-JUN-1999;	99US-0139750.	PR	23-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	99US-0139763.	PR	25-AUG-1999;	99US-0150566.
PR	21-JUN-1999;	99US-0139817.	PR	26-AUG-1999;	99US-0150884.
PR	22-JUN-1999;	99US-0139899.	PR	27-AUG-1999;	99US-0151065.
PR	23-JUN-1999;	99US-0140353.	PR	27-AUG-1999;	99US-0151066.
PR	23-JUN-1999;	99US-0140354.	PR	27-AUG-1999;	99US-0151080.
PR	24-JUN-1999;	99US-0140695.	PR	30-AUG-1999;	99US-0151303.
PR	28-JUN-1999;	99US-0140823.	PR	31-AUG-1999;	99US-0151438.
PR	30-JUN-1999;	99US-0141287.	PR	01-SEP-1999;	99US-0151930.
PR	01-JUL-1999;	99US-0141842.	PR	07-SEP-1999;	99US-0152363.
PR	01-JUL-1999;	99US-0142154.	PR	10-SEP-1999;	99US-0153070.
PR	02-JUL-1999;	99US-0142055.	PR	13-SEP-1999;	99US-0153758.
PR	06-JUL-1999;	99US-0142390.	PR	15-SEP-1999;	99US-0154018.
PR	08-JUL-1999;	99US-0142803.	PR	16-SEP-1999;	99US-0154039.
			PR	20-SEP-1999;	99US-0154779.
			PR	22-SEP-1999;	99US-0155139.
			PR	23-SEP-1999;	99US-0155486.
			PR	24-SEP-1999;	99US-0155659.

PR	28-SEP-1999;	99US-0156458.	
PR	29-SEP-1999;	99US-0156596.	
PR	04-OCT-1999;	99US-0157117.	
PR	05-OCT-1999;	99US-0157753.	
PR	06-OCT-1999;	99US-0157865.	
PR	07-OCT-1999;	99US-0158029.	
PR	08-OCT-1999;	99US-0158232.	
PR	12-OCT-1999;	99US-0158369.	
PR	13-OCT-1999;	99US-0159293.	
PR	13-OCT-1999;	99US-0159294.	
PR	13-OCT-1999;	99US-0159295.	
PR	14-OCT-1999;	99US-0159329.	
PR	14-OCT-1999;	99US-0159330.	
PR	14-OCT-1999;	99US-0159331.	
PR	14-OCT-1999;	99US-0159637.	
PR	14-OCT-1999;	99US-0159638.	
PR	18-OCT-1999;	99US-0159584.	
PR	21-OCT-1999;	99US-0160741.	
PR	21-OCT-1999;	99US-0160767.	
PR	21-OCT-1999;	99US-0160768.	
PR	21-OCT-1999;	99US-0160770.	
PR	21-OCT-1999;	99US-0160814.	
PR	21-OCT-1999;	99US-0160815.	
PR	22-OCT-1999;	99US-0160980.	
PR	22-OCT-1999;	99US-0160981.	
PR	22-OCT-1999;	99US-0160989.	
PR	25-OCT-1999;	99US-0161404.	
PR	25-OCT-1999;	99US-0161405.	
PR	25-OCT-1999;	99US-0161406.	
PR	26-OCT-1999;	99US-0161359.	
PR	26-OCT-1999;	99US-0161360.	
PR	26-OCT-1999;	99US-0161361.	
PR	28-OCT-1999;	99US-0161920.	
PR	28-OCT-1999;	99US-0161992.	
PR	28-OCT-1999;	99US-0161993.	
PR	29-OCT-1999;	99US-0162142.	
Query Match 2.5%; Score 39.2; DB 21; Length 1554;			
Best Local Similarity 51.7%; Pred. No. 0.11;			
Matches 89; Conservative 0; Mismatches 83; Indels 0; Gaps 0;			
QY	176	ctgggtatctcaggcaggtctccagctcgccctactactagcggggggcccgtaactctctctt 235	
Db	149	ctgggtatcaccaccactacaacacggcgttaagattcggcggaactgtcaactctggtct 208	
QY	236	acgggcactcctaatacgtactctctcgctctacatctgcatacgtttctcattagccgaactga 295	
Db	209	acggatggttctctgcgggtctcttcacaatgtgcgttggttatctatggtccgagatct 268	
QY	296	ccagcgtctaccgcagctgcgggtggccaatatcatcttttgcgtcgatcctgcg 347	
Db	269	gctctcttaccctacctccggtggtctctactactggaagtgtatgctgcgc 320	
RESULT 6			
AAC49990			
ID	AAC49990 standard; DNA; 1664 BP.		
XX			
AC	AAC49990;		
XX			
DT	18-OCT-2000 (first entry)		
XX			
DE	Arabidopsis thaliana DNA fragment SEQ ID NO: 63189.		
XX			
KW	Hybridisation assay; genetic mapping; gene expression control;		
KW	protein identification; signal transduction pathway;		
KW	metabolic pathway; promoter; termination sequence; ss.		
XX			
OS	Arabidopsis thaliana.		
XX			
PN	EP1033405-A2.		
XX			
PD	06-SEP-2000.		

XX	25-FEB-2000;	2000EP-0301439.	
PF			
XX			
PR	25-FEB-1999;	99US-0121825.	
PR	05-MAR-1999;	99US-0123180.	
PR	09-MAR-1999;	99US-0123548.	
PR	23-MAR-1999;	99US-0125788.	
PR	25-MAR-1999;	99US-0126264.	
PR	29-MAR-1999;	99US-0126785.	
PR	01-APR-1999;	99US-0127462.	
PR	06-APR-1999;	99US-0128234.	
PR	08-APR-1999;	99US-0128714.	
PR	16-APR-1999;	99US-0129845.	
PR	19-APR-1999;	99US-0130077.	
PR	21-APR-1999;	99US-0130449.	
PR	23-APR-1999;	99US-0130510.	
PR	23-APR-1999;	99US-0130891.	
PR	28-APR-1999;	99US-0131449.	
PR	30-APR-1999;	99US-0132048.	
PR	30-APR-1999;	99US-0132407.	
PR	04-MAY-1999;	99US-0132484.	
PR	05-MAY-1999;	99US-0132485.	
PR	06-MAY-1999;	99US-0132486.	
PR	06-MAY-1999;	99US-0132487.	
PR	07-MAY-1999;	99US-0132863.	
PR	11-MAY-1999;	99US-0134256.	
PR	14-MAY-1999;	99US-0134218.	
PR	14-MAY-1999;	99US-0134219.	
PR	14-MAY-1999;	99US-0134221.	
PR	14-MAY-1999;	99US-0134370.	
PR	18-MAY-1999;	99US-0134768.	
PR	19-MAY-1999;	99US-0134941.	
PR	20-MAY-1999;	99US-0135124.	
PR	21-MAY-1999;	99US-0135353.	
PR	24-MAY-1999;	99US-0135629.	
PR	25-MAY-1999;	99US-0136021.	
PR	27-MAY-1999;	99US-0136392.	
PR	28-MAY-1999;	99US-0136782.	
PR	01-JUN-1999;	99US-0137222.	
PR	03-JUN-1999;	99US-0137528.	
PR	04-JUN-1999;	99US-0137502.	
PR	07-JUN-1999;	99US-0137724.	
PR	08-JUN-1999;	99US-0138094.	
PR	10-JUN-1999;	99US-0138540.	
PR	10-JUN-1999;	99US-0138847.	
PR	14-JUN-1999;	99US-0139119.	
PR	16-JUN-1999;	99US-0139452.	
PR	16-JUN-1999;	99US-0139453.	
PR	17-JUN-1999;	99US-0139492.	
PR	18-JUN-1999;	99US-0139454.	
PR	18-JUN-1999;	99US-0139455.	
PR	18-JUN-1999;	99US-0139456.	
PR	18-JUN-1999;	99US-0139457.	
PR	18-JUN-1999;	99US-0139458.	
PR	18-JUN-1999;	99US-0139459.	
PR	18-JUN-1999;	99US-0139460.	
PR	18-JUN-1999;	99US-0139461.	
PR	18-JUN-1999;	99US-0139462.	
PR	18-JUN-1999;	99US-0139463.	
PR	18-JUN-1999;	99US-0139750.	
PR	18-JUN-1999;	99US-0139763.	
PR	21-JUN-1999;	99US-0139817.	
PR	22-JUN-1999;	99US-0139899.	
PR	23-JUN-1999;	99US-0140353.	
PR	23-JUN-1999;	99US-0140354.	
PR	24-JUN-1999;	99US-0140695.	
PR	28-JUN-1999;	99US-0140823.	
PR	29-JUN-1999;	99US-0140991.	
PR	30-JUN-1999;	99US-0141287.	
PR	01-JUL-1999;	99US-0141842.	
PR	01-JUL-1999;	99US-0142154.	
PR	02-JUL-1999;	99US-0142055.	
PR	06-JUL-1999;	99US-0142390.	

PR	08-JUL-1999;	99US-0142803
PR	09-JUL-1999;	99US-0142920
PR	12-JUL-1999;	99US-0142977
PR	13-JUL-1999;	99US-0143542
PR	14-JUL-1999;	99US-0143624
PR	15-JUL-1999;	99US-0144005
PR	16-JUL-1999;	99US-0144085
PR	16-JUL-1999;	99US-0144086
PR	19-JUL-1999;	99US-0144325
PR	19-JUL-1999;	99US-0144331
PR	19-JUL-1999;	99US-0144332
PR	19-JUL-1999;	99US-0144333
PR	19-JUL-1999;	99US-0144334
PR	19-JUL-1999;	99US-0144335
PR	20-JUL-1999;	99US-0144352
PR	20-JUL-1999;	99US-0144632
PR	20-JUL-1999;	99US-0144884
PR	21-JUL-1999;	99US-0144814
PR	21-JUL-1999;	99US-0145086
PR	21-JUL-1999;	99US-0145088
PR	22-JUL-1999;	99US-0145085
PR	22-JUL-1999;	99US-0145087
PR	22-JUL-1999;	99US-0145089
PR	22-JUL-1999;	99US-0145192
PR	23-JUL-1999;	99US-0145145
PR	23-JUL-1999;	99US-0145218
PR	23-JUL-1999;	99US-0145224
PR	26-JUL-1999;	99US-0145276
PR	27-JUL-1999;	99US-0145913
PR	27-JUL-1999;	99US-0145918
PR	27-JUL-1999;	99US-0145919
PR	28-JUL-1999;	99US-0145951
PR	02-AUG-1999;	99US-0146386
PR	02-AUG-1999;	99US-0146389
PR	02-AUG-1999;	99US-0146388
PR	03-AUG-1999;	99US-0147038
PR	04-AUG-1999;	99US-0147204
PR	04-AUG-1999;	99US-0147302
PR	05-AUG-1999;	99US-0147192
PR	05-AUG-1999;	99US-0147260
PR	06-AUG-1999;	99US-0147303
PR	06-AUG-1999;	99US-0147416
PR	09-AUG-1999;	99US-0147493
PR	09-AUG-1999;	99US-0147935
PR	10-AUG-1999;	99US-0148171
PR	11-AUG-1999;	99US-0148319
PR	12-AUG-1999;	99US-0148341
PR	13-AUG-1999;	99US-0148565
PR	13-AUG-1999;	99US-0148684
PR	16-AUG-1999;	99US-0149368
PR	17-AUG-1999;	99US-0149175
PR	18-AUG-1999;	99US-0149426
PR	20-AUG-1999;	99US-0149722
PR	20-AUG-1999;	99US-0149723
PR	20-AUG-1999;	99US-0149929
PR	23-AUG-1999;	99US-0149902
PR	23-AUG-1999;	99US-0149930
PR	25-AUG-1999;	99US-0150566
PR	26-AUG-1999;	99US-0150884
PR	27-AUG-1999;	99US-0151065
PR	27-AUG-1999;	99US-0151066
PR	27-AUG-1999;	99US-0151080
PR	30-AUG-1999;	99US-0151303
PR	31-AUG-1999;	99US-0151438
PR	01-SEP-1999;	99US-0151930
PR	07-SEP-1999;	99US-0152363
PR	10-SEP-1999;	99US-0153070
PR	13-SEP-1999;	99US-0153758
PR	15-SEP-1999;	99US-0154018
PR	16-SEP-1999;	99US-0154039
PR	20-SEP-1999;	99US-0154779
PR	22-SEP-1999;	99US-0155139
PR	23-SEP-1999;	99US-0155486

PR	24-SEP-1999;	99US-01555659
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

Query Match 2.5%; Score 39.2; DB 21; Length 1664;
Best Local Similarity 51.7%; Pred. No. 0.11;
Matches 89; Conservative 0; Mismatches 83; Indels 0;

QY	176	ctggatatcaggcaggtctccagctcgccctactagcgggggggcccgcctcactctccttt	235
Db	184	ctggatatcaccaccacctacacaccggcttaagattcggcggcactgctcactctgggtct	243
QY	236	acggcattcctaatacagtagtctctgctctacatctgcatactcgcttctctcattagccgaactga	295
Db	244	acggatgggttcctggcgggctccttcacaatgtgcgttgggttatctctatgcccgaagatct	303
QY	296	ccagcgtctataccgcgactgccgggtggccaatatctttgcgctcgatccctggc	347
Db	304	gctcctctataccctacacctccgggtggtctctactactcgaagtgcctatgctcgc	355

RESULT	7	
AAV46335/C		
ID	AAV46335	standard; cDNA; 5839 BP.
XX		
AC	AAV46335;	
XX		
DT	20-NOV-1998	(first entry)
XX		
DE	Hydra head activator binding protein cDNA.	
XX		
KW	Head activator binding protein; hydra; screening assay; agonist;	
KW	antagonist; treatment; neurodegenerative disease; ss.	
XX		
OS	Chlorohydra viridissima.	
XX		
Key	Location/Qualifiers	
FT	481..5466	
FT	/+tag= a	
FT		

FT sig_peptide 481..567
FT mat_peptide /*tag= b
FT 568..5463
FT /*tag= C
FT /product= "head activator binding protein"

PN DE19808258-A1.

XX 03-SEP-1998.

XX 27-FEB-1998; 98DE-1008258.

PR 10-JUL-1997; 97DE-1029490.

XX 28-FEB-1997; 97DE-1008134.

XX (EVOT-) EVOTEC BIOSYSTEMS GMBH.

PA Hampe W, Schaller C;

PI WPI; 1998-468562/41.

XX P-PSDB; AAW76041.

DR Hydra head activator binding protein - useful in screening assay for

XX agonists or antagonists

PS Claim 18; Page 11-14; 21pp; German.

XX This sequence encodes a head activator binding protein isolated from
CC Chlorohydra viridissima by chromatographic methods. The protein is used
CC in a screening assay for agonists or antagonists comprising determining
CC the binding properties of candidate substances with the protein. The
CC assay can be used to screen for substances useful for treating
CC neurodegenerative diseases.

XX Sequence 5839 BP; 1856 A; 978 C; 1106 G; 1899 T; 0 other;

Query Match 2.5%; Score 39.2; DB 19; Length 5839;
Best Local Similarity 56.1%; Pred. No. 0.23;
Matches 74; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 1365 ggtggcttcacgltccgtcaccactgtgtttttcagcttccactgacogtgactacggc 1424

Db 3883 GTTTTTCATCACTACCATCCCAATCATTTTCACCATCACAAATGCCATGACAAACTTG 3824

QY 1425 cgcgtcaaccatgaattacacaaagtgcgattataggcgttgcaacttgctcttggtgtctt 1484

Db 3823 GTAAACAACGATTATTGTCACATGTGAATGGTTAGCGTGCACCTTTACTTTATTGCATA 3764

QY 1485 gaactgggtcgt 1496

Db 3763 TATTTTGTCTT 3752

RESULT 8
AAA10595/c
ID AAA10595 standard; DNA; 6741 BP.
XX
AC AAA10595;
XX
DT 29-JUN-2000 (first entry)
XX
DE Gene encoding a subunit of cellulose synthase.
XX
KW Cellulose synthase; cellulose production; increase yield; ds.
XX
OS Vigna angularis.
XX
PN JP2000060568-A.
XX
PD 29-FEB-2000.
XX
PF 26-AUG-1998; 98JP-0239998.

XX 26-AUG-1998; 98JP-0239998.
PR (MIZU/) MIZUNO K.
PA (OJIP) OJI PAPER CO.
XX
XX WPI; 2000-342371/30.
DR P-PSDB; AAY85180.
XX
PT A gene encoding a cellulose synthetic equipment - for the improvement
PT in the amount of cellulose synthesised in a plant body
XX
XX Claim 2; Page 22-31; 32pp; Japanese.
XX
CC This sequence represents a gene encoding a subunit of the cellulose
CC synthase complex of Vigna angularis. The invention relates to subunits of
CC cellulose synthetic equipment, that can be used to increase the amount of
CC cellulose synthesised by a plant. The proteins and genes encoding them
CC can also be used to improve the properties of the cellulose being
CC produced by a plant.
XX
SQ Sequence 6741 BP; 1712 A; 870 C; 1468 G; 1374 T; 1317 other;

Query Match 2.4%; Score 38.2; DB 21; Length 6741;
Best Local Similarity 19.6%; Pred. No. 0.52;
Matches 46; Conservative 88; Mismatches 101; Indels 0; Gaps 0;

QY 126 tggcaccatcacctcgtgtccttggccttggatttgcaacagttgggctggtatctc 185

Db 5880 TBDRCYRARCSTRYSGRCYSYSTYDATTTBSRSGYSNPRCDYDAYSDYDASTDTS 5821

QY 186 aggcagtcctccagctcgcctactagcgggggggcccgcctcctcttaccggcactc 245

Db 5820 RSDSTTTCTSTNSTSRDNSTRCTTRCYSSRCYTSYSTCTSDYSYRAAKCYTSRNST 5761

QY 246 aatcagtcactcgtctacatctgcctgttctcattagcgcgaactgaccagcgtctc 305

Db 5760 NSTYYDADNSTTBSSTTCYTTTSTNCTNSTYSYRANSTTBYRARCYSCTCYT 5701

QY 306 cccgactgccgtggccaatatcatcttgcgctgcactcctggcaccacaaatcaatc 360

Db 5700 YSYDASTCYTSRGTBCYTTBYSTBTBNCRNCNCYSDSYRANCYSSTYDATBNSTY 5646

RESULT 9
AAQ92829
ID AAQ92829 standard; DNA; 3510 BP.
XX
AC AAQ92829;
XX
DT 21-FEB-1996 (first entry)
XX
DE Mitochondrial cytochrome P450/NADPH P450 reductase fusion DNA.

XX
KW Mitochondrial; cytochrome P450; NADPH; reductase; mature c25;
KW industrial waste; biological oxidation; fusion DNA; ds.
XX
OS Homo sapiens.

XX
FH Key Location/Qualifiers
FT CDS 1..3510
FT /*tag= a

XX JP07147975-A.

XX 13-JUN-1995.

XX 29-NOV-1993; 93JP-0298279.

XX 29-NOV-1993; 93JP-0298279.

XX (SUMO) SUMITOMO CHEM CO LTD.

```
XX WPI; 1995-242753/32.
DR P-PSDB; AAR76544.
XX
PT Fusion protein comprising mitochondrial cytochrome P450 and
PT NADPH-cytochrome P450 reductase fragments - may be used in oxidation
PT processes, particularly treatment of industrial waste
XX
PS Claim 2; Pages 5-10; 11pp; Japanese.
XX
CC AAQ92829 encodes AAR76544 a fusion protein comprising mature
CC mitochondrial cytochrome P450 c25 at its N-terminus, and NADPH
CC cytochrome P450 reductase (which catalyses the reduction of
CC microsomal type cytochrome P450) at its C-terminus. The fusion
CC protein may be used for the biological oxidation of useful
CC materials, and for the treatment of industrial wastes.
XX
SQ Sequence 3510 BP; 889 A; 911 C; 838 G; 872 T; 0 other;

Query Match      2.4%; Score 37.8; DB 16; Length 3510;
Best Local Similarity 60.0%; Pred. No. 0.48;
Matches 63; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 6 ctccagaccaaagtgtatcgccgagaaaggcgggacaaaggcagacaaacgaacacaga 65
   ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 53 ctgcagccctccgggatacagagacgagacgaggggtccaggaaacaggtcaagaccgcgc 112
   ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 66 gacggcggcgccaggtgtgctccgagtcctgaacgttcctct 110
   ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 113 gcctgcggagctctggcggagcttccgggaccgcggaacgcgtacgct 157
   ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 10
AAC38835
ID AAC38835 standard; DNA; 1686 BP.
XX
AC AAC38835;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 22404.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
```


PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145192.
PR	23-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145218.
PR	23-JUL-1999;	99US-0145224.
PR	26-JUL-1999;	99US-0145276.
PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.
PR	27-JUL-1999;	99US-0145919.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.
Query Match 2.4%; Score 37.6; DB 21; Length 1686;		
Best Local Similarity 51.2%; Pred. No. 0.36; Mismatches 84; Indels 0; Gaps 0;		
Matches 88; Conservative 0;		
QY	176	ctggtatctcaggcagtcctccagtcgcccctactagcggggggcccgctcaactctcctt 235
DB	186	ctggtatcaccaccacctacacacgggttaagattcggcgccactgtcactctgtct 245
QY	236	acggcatcctaatacagtcactctctgtatctacatctgcacgtcttctcattagccgaactga 295
DB	246	acggatggttcctgcccggctcctctcaaatgtcgttgggttatctatggtcgtgatctct 305
QY	296	ccagcgtctacccgactgcgggtgggccaataatcattttgcgtcgatccctggc 347
DB	306	gtcctcttaccctacccctccgggtgtctctctactactgagtgatgctatgctgc 357
RESULT 11		
AAA10594/c		
ID	AAA10594 standard; DNA: 10732 BP.	
XX		
AC	AAA10594;	
DT	29-JUN-2000 (first entry)	
XX	Gene encoding a subunit of cellulose synthase.	
DE	Cellulose synthase; cellulose production; increase yield; ds.	
XX	Vigna angularis.	
KW	JP2000060568-A.	
XX	29-FEB-2000.	
XX	26-AUG-1998; 98JP-0239998.	
PF	26-AUG-1998; 98JP-0239998.	
XX	(MIZU/) MIZUNO K.	
PA	(OJIP) OJI PAPER CO.	
XX	WPI; 2000-342371/30.	
DR	P-PSDB; AAY85179.	
XX	A gene encoding a cellulose synthetic equipment - for the improvement	
PT	in the amount of cellulose synthesised in a plant body	
XX	Claim 2; Page 14-21; 32pp; Japanese.	
XX	This sequence represents a gene encoding a subunit of the cellulose	
CC	synthase complex of Vigna angularis. The invention relates to subunits of	
CC	cellulose synthetic equipment, that can be used to increase the amount of	
CC	cellulose synthesised by a plant. The proteins and genes encoding them	
CC	can also be used to improve the properties of the cellulose being	
CC	produced by a plant.	


```
FT intron 93591..97475
FT /*tag= x
FT /label= K
FT 97476..97960
FT /*tag= y
FT /label= L
FT 988..1006
FT /*tag= z
FT /bound_moiety= "Primer B2"
FT /note= "Amplification of amplicon 99-20508"
FT complement (1509..1529)
FT /*tag= aa
FT /bound_moiety= "Primer C2"
FT /note= "Amplification of amplicon 99-20508"
FT 5039..5056
FT /*tag= ab
FT /bound_moiety= "Primer B3"
FT /note= "Amplification of amplicon 99-20469"
FT complement (5534..5554)
FT /*tag= ac
FT /bound_moiety= "Primer C3"
FT /note= "Amplification of amplicon 99-20469"
FT 5997..6015
FT /*tag= ad
FT /bound_moiety= "Primer B4"
FT /note= "Amplification of amplicon 5-254"
FT complement (6332..6350)
FT /*tag= ae
FT /bound_moiety= "Primer C4"
FT /note= "Amplification of amplicon 5-254"
FT 14371..14390
FT /*tag= af
FT /bound_moiety= "Primer B5"
FT /note= "Amplification of amplicon 5-257"
FT complement (14798..14817)
FT /*tag= ag
FT /bound_moiety= "Primer C5"
FT /note= "Amplification of amplicon 5-257"
FT 18751..18771
FT /*tag= ah
FT /bound_moiety= "Primer B6"
FT /note= "Amplification of amplicon 99-20511"
FT complement (19198..19217)
FT /*tag= ai
FT /bound_moiety= "Primer C6"
FT /note= "Amplification of amplicon 99-20511"
FT 19605..19625
FT /*tag= aj
FT /bound_moiety= "Primer B7"
FT /note= "Amplification of amplicon 99-20510"
FT complement (19986..20005)
FT /*tag= ak
FT /bound_moiety= "Primer C7"
FT /note= "Amplification of amplicon 99-20510"
FT 29529..29547
FT /*tag= al
FT /bound_moiety= "Primer B8"
FT /note= "Amplification of amplicon 99-20504"
FT complement (30041..30061)
FT /*tag= am
FT /bound_moiety= "Primer C8"
FT /note= "Amplification of amplicon 99-20504"
FT 42268..42287
FT /*tag= an
FT /bound_moiety= "Primer B9"
FT /note= "Amplification of amplicon 99-20493"
FT complement (42732..42752)
FT /*tag= ao
FT /bound_moiety= "Primer C9"
FT /note= "Amplification of amplicon 99-20493"
FT 69026..69046
FT /*tag= ap
FT /bound_moiety= "Primer B10"
```

```
FT primer_bind
FT /note= "Amplification of amplicon 99-20499"
FT complement (69525..69543)
FT /*tag= aq
FT /bound_moiety= "Primer C10"
FT /note= "Amplification of amplicon 99-20499"
FT 76323..76343
FT /*tag= ar
FT /bound_moiety= "Primer B11"
FT /note= "Amplification of amplicon 99-20473"
FT complement (76771..76790)
FT /*tag= as
FT /bound_moiety= "Primer C11"
FT /note= "Amplification of amplicon 99-20473"
FT 78292..78309
FT /*tag= at
FT /bound_moiety= "Primer B12"
FT /note= "Amplification of amplicon 5-249"
FT complement (78704..78721)
FT /*tag= au
FT /bound_moiety= "Primer C12"
FT /note= "Amplification of amplicon 5-249"
FT 81893..81912
FT /*tag= av
FT /bound_moiety= "Primer B13"
FT /note= "Amplification of amplicon 99-20485"
FT complement (82353..82372)
FT /*tag= aw
FT /bound_moiety= "Primer C13"
FT /note= "Amplification of amplicon 99-20485"
FT 84392..84412
FT /*tag= ax
FT /bound_moiety= "Primer B14"
FT /note= "Amplification of amplicon 99-20481"
FT complement (84909..84929)
FT /*tag= ay
FT /bound_moiety= "Primer C14"
FT /note= "Amplification of amplicon 99-20481"
FT 89746..89765
FT /*tag= az
FT /bound_moiety= "Primer B15"
FT /note= "Amplification of amplicon 99-20480"
FT complement (90179..90198)
FT /*tag= ba
FT /bound_moiety= "Primer C15"
FT /note= "Amplification of amplicon 99-20480"
FT 1431..1455
FT /*tag= bb
FT /bound_moiety= "Probe P2"
```

Query Match 2.3%; Score 36.6; DB 21; Length 99960;
Best Local Similarity 54.0%; Pred. No. 8.4;
Matches 75; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

```
Qy 359 tcaatcggagcatttcatacgtgtgcggactcgttgcgttgcgttcattgcgtatcg 418
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 24794 tgaacggaggctgccctggttctgtgcactgcctcgtgagtgtgtggaaggctggg 24853

Qy 419 gaagctcagtgaccatgatacctgcgtcaacagatcccgcgctgataccgctatagtc 478
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 24854 gaagccaaagtctccatggtgcctccatcagggaacctgcagctggaggcagccagagg 24913

Qy 479 acacatactcccaggattc 497
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 24914 ccacagggttgtagcattc 24932
```

RESULT 13
AAQ14834
ID AAQ14834 standard; DNA; 1900 BP.
XX
AC AAQ14834;
XX
DT 10-FEB-1992 (first entry)

```
XX Cytochrome P450C25 gene.
DE Vitamin D3; ss.
XX Rattus rattus.
XX Key Location/Qualifiers
FT CDS 59..1660
FT sig_peptide /*tag= a
FT mat_peptide 59..154
FT 155..1657 /*tag= b
FT /*tag= c

XX JP03232493-A.
XX 16-OCT-1991.
XX 06-FEB-1990; 90JP-0027711.
XX 06-FEB-1990; 90JP-0027711.
XX (SUMO ) SUMITOMO CHEM IND KK.
XX WPI; 1991-349005/48.
XX P-PSDB; AARI15057.
XX Cytochrome P450C25 gene expressed by microbe cells - is used to
PT produce enzyme for mfr. of active vitamin-D3, etc.
XX Claim 1; Fig 1; 10pp; Japanese.
XX The sequence is that of the cytochrome P450C25 gene which was obt'd.
CC from rat liver. It can be recombinantly expressed in microorganisms
CC in sufficient amounts to be used in the process of manufacturing
CC active vitamin D3.
XX Sequence 1900 BP; 451 A; 552 C; 476 G; 421 T; 0 other;

Query Match 2.3%; Score 36.2; DB 12; Length 1900;
Best Local Similarity 59.0%; Pred. No. 1.1;
Matches 62; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 6 ctccagacaaagtggatcgcgagagaagggcggaagcagagcagacgaacacaga 65
Db 162 ctgcagccctccgcgatcagagacgagcagggggtccaggaacaggtcaagaccgacgc 221

QY 66 gacggcgggcagggtggtggtccgagtcctgaacgttctct 110
Db 222 gcctgcggagctcggcgaggttcgggacccggacggacgctacgct 266

RESULT 14
AAF07981
ID AAF07981 standard; cDNA; 2036 BP.
XX
AC AAF07981;
XX
DT 13-MAR-2001 (first entry)
DE Fusarium venenatum EST SEQ ID NO:504.
XX
KW Multiple gene expression; filamentous fungal cell; EST;
KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW culture condition; environmental stress; spore morphogenesis;
KW metabolic pathway engineering; catabolic pathway engineering; ss.
XX
OS Fusarium venenatum.
XX
PN WO200056762-A2.
```

```
XX 28-SEP-2000.
XX
XX 22-MAR-2000; 2000WO-US07781.
XX
XX 22-MAR-1999; 99US-0273623.
XX
XX (NOVO ) NOVO NORDISK BIOTECH INC.
XX (NOVO ) NOVO NORDISK AS.
XX
XX Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX WPI; 2000-594572/56.
XX
XX Monitoring differential expression of genes in filamentous fungal cells
XX uses fluorescence-labeled nucleic acids isolated from the cells and a
XX substrate of expressed sequence tags -
XX
XX Claim 86; Page 578; 3161pp; English.
XX
XX The present invention describes a method for monitoring differential
XX expression of genes in a first filamentous fungal (FF) cell relative to
XX expression of the same genes in one or more second filamentous fungal
XX cells. The method uses fluorescence-labeled nucleic acids isolated from
XX the FF cells and a substrate of expressed sequence tags (EST). The ESTs
XX are used in the methods for monitoring differential expression of genes
XX in a first filamentous fungal (FF) cell relative to expression of the
XX same genes in one or more second filamentous fungal cells. Monitoring
XX the global expression of genes from FF cells allows the production
XX potential of the microorganisms to be improved. New genes may be
XX discovered, possible functions of unknown open reading frames can be
XX identified and gene copy number variation and stability can be
XX monitored. The expression of genes can be used to study how FF cells
XX adapt to changes in culture conditions, environmental stress, spore
XX morphogenesis, recombination, metabolic or catabolic pathway
XX engineering. Using ESTs provides several advantages over genomic or
XX random cDNA clones including elimination of redundancy as one spot on an
XX array equals one gene or open reading frame, and organisation of the
XX microarrays based on function of the gene products to facilitate
XX analysis of the results. AAF07478 to AAF11247 represents ESTs from
XX Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
XX niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
XX AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
XX all specifically claimed in the present invention.
XX
XX Sequence 2036 BP; 399 A; 582 C; 523 G; 521 T; 11 other;

Query Match 2.2%; Score 35.4; DB 21; Length 2036;
Best Local Similarity 47.5%; Pred. No. 2.1;
Matches 105; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

QY 130 accatcaccatcggtccttggtggtggtggtggtggtggtggtggtggtggtggt 189
Db 236 actttccaggtcgcttttccttccttccttccttccttccttccttccttccttc 295

QY 190 agtctccagctcgccctactagcgggggggggggggggggggggggggggggggg 249
Db 296 acctgggttaccctctgattggcggtggccctgtcaacgttatctctgggtggtcgc 355

QY 250 agtactctcgctctacatctgcatcgcttttctcattagccgaactgaccagcgct 309
Db 356 gtctctctcattattgtctgtgctgctgtctctctctctctctctctctctctct 415

QY 310 actgcccgggtggccaataatcatttttgcgtcgatccttggcacc 350
Db 416 acaactggaagtgtctactactaccaagcttttcattgctctccc 456

RESULT 15
AAH24065/C
ID AAH24065 standard; DNA; 4590 BP.
XX
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2002, 23:04:52 ; Search time 7811.87 Seconds
(without alignments)
4227.166 Million cell updates/sec

Title: US-09-882-694A-7
Perfect score: 1578
Sequence: 1 atggactccagaccagtggtg.....aatttcaagttgggccaatga 1578

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
-----	-----	-----	-----	-----

1	1555.6	98.6	1578	6	AX076848	Sequence
2	1555.6	98.6	1764	6	AX076847	Sequence
3	66.4	4.2	2475	8	SCYGL077C	S. cerevisia
4	66.4	4.2	2725	8	YSCCTR	Saccharocyc
5	58	3.7	7218	6	I66494	Sequence 14
6	52.4	3.3	3724	8	YSCSEQ	Saccharomyc
7	47.4	3.0	125020	9	AF429315	Homo sapi
8	44.8	2.8	27893	8	SPAC1039	S. pombe c
9	41.2	2.6	147203	8	AP003687	Oryza sat
10	41	2.6	125020	9	AF429315	Homo sapi
11	40	2.5	1996	8	AF053231	Neurospor
12	39.2	2.5	720	8	CNS01D5D	Botrytis
13	39.2	2.5	5359	3	AF092920	Chlorohyd
14	39.2	2.5	84957	8	AC006200	Arabidops
15	38.2	2.4	6741	6	E32987	Gene encodi
16	38.2	2.4	286550	1	SME591785	Sinorhizo
17	37.8	2.4	1201	10	RNCYT27S1	Rattus norv
18	37.8	2.4	1877	10	RATMTCYPCH	Rat mitocho
19	37.8	2.4	1900	6	E03144	CDNA encodi
20	37.8	2.4	1900	10	RNCYP27	Rattus norv
21	37.8	2.4	2300	10	RATP45OC	Sprague-Daw
22	37.8	2.4	3510	23	E09419	Synthetic n
23	37.6	2.4	1726	8	AF019637	Arabidops
24	37.4	2.4	981	11	CNS06JH1	T3 end of
25	37.4	2.4	110000	2	AC013622_0	Mus muscu
26	37.4	2.4	110000	2	AC021632_0	Mus muscu
27	37.2	2.4	1246	6	AX164174	Sequence
28	37.2	2.4	1891	8	SCUGA4	S. cerevisia
29	37.2	2.4	5794	8	SCYDL210W	S. cerevisia
30	37.2	2.4	39090	8	CS39KBCIV	S. cerevisia
31	36.8	2.3	12062	1	SCD17	Streptomy
32	36.8	2.3	329709	1	AP002997	Mesorhizo
33	36.6	2.3	10732	6	E32986	Gene encodi
34	36.6	2.3	154344	9	AC073308	Homo sapi
35	36.4	2.3	11514	1	AE002028	Deinococc
36	36.4	2.3	69634	2	AC101748	Mus muscu
37	36.4	2.3	105213	2	AL158217	Homo sapi
38	36.4	2.3	203188	2	AC104323	Mus muscu
39	36	2.3	183876	9	AC006390	Homo sapi
40	35.8	2.3	289	6	AR162089	Sequence
41	35.8	2.3	289	6	AR166614	Sequence
42	35.8	2.3	13848	1	AE005104	Halobacte
43	35.8	2.3	62299	2	AC100224	Mus muscu
44	35.8	2.3	145911	1	AP003014	Mesorhizo
45	35.8	2.3	166335	2	AC073799	Mus muscu

ALIGNMENTS

RESULT	1
AX076848	
LOCUS	AX076848
DEFINITION	Sequence 7 from Patent WO0105980.
ACCESSION	AX076848
VERSION	AX076848.1
KEYWORDS	GI:13121521
SOURCE	Exophiala spinifera.
ORGANISM	Exophiala spinifera
REFERENCE	1 (bases 1 to 1578)
AUTHORS	Duval, J.P., Maddox, J., Gilliam, J., Folkerts, O. and Crasta, O.R.
TITLE	Compositions and methods for fumonisins detoxification
JOURNAL	Patent: WO 0105980-A 7 25-JAN-2001;
FEATURES	Pioneer Hi-Bred International Inc. (US) ; Curagen Corporation (US)
source	Location/Qualifiers
	1..1578
	/organism="Exophiala spinifera"
	/db_xref="taxon:91928"
	/note="permease, coding sequence"
BASE COUNT	324 a 446 c 385 g 423 t

QY 1 atggaactccagaccagaatggatagcggcgagaaaggcgggacaaaggcagacaaacgaagaac 60
Db 5 ATGGACTCCAGACCAAGTGGATACGGCTTGAAAGGCGGGACAAAGGCAGACAACGAGAAC 64
QY 61 acagagacggcgggcgaggtggtgctgcagagtcctcgaacgttccctctggaagaagaa 120
Db 65 ACAGAGACGGCGGCGCAGGTGGTGGTCCGAGTCCCTGAACGTTCTCTGGAGAGAAA 124
QY 121 caatttggcaccatcaccatcgtgtcccttggccttctgtgatttgcacagtttggcgtgt 180
Db 125 CAATTGGCACCATCACCATCGTGTCTTGGCCTTTGTGATTGCAACAGTTGGGCTGGT 184
QY 181 atctcaggcagctctccagctcgccctactagcggggggggcccgctcactctcctttacggc 240
Db 185 ATCTAGGCAGTCTCCAGCTCGCCCTACTAGCGGGGGGGCCCGTCACTCTCCTTACGGC 244
QY 241 atcctaatacgtactctcgtctacatctgcatcgtcttctcattagccgaactgaccagc 300
Db 245 ATCCTAATCAGTACTCTCGTCTACATCTGCATCGCTTTCTCAATTAGCCGAACTGACCAGC 304
QY 301 gtctacccgactgcgggtggcccaatatcatatttgcgtcgatcctctggcaccacaaatac 360
Db 305 GTCTACCCGACTGCCGGTGGCCAAATATCATTTTGGCTCGATCCYGGCACCAAAATCAATC 364
QY 361 aatcgagagcatttcaatcagtgctgaggactcgtgtcgttgcgttcttcatggatcgtatcgga 420
Db 365 AATCGGAGCATTTCAATACGTGTGGGACTCGTGTCTGCTTTCATGGATCGCTATCGGA 424
QY 421 agctcagtgaccatgataacctgctcaacagatccccggcgctgatagccgcctatagtcaac 480
Db 425 AGCTCAGTGACCATGATACCTGCTCAACAGATCCCGCGCTGATAGACCGCCTATAGTCAAC 484
QY 481 acatactcccaggattcgtggcatgtcttccctcatctacgagggagtcgcgtggtgtg 540
Db 485 ACATACTCCAGGATTCGTGGCATGTCTTCCCTCATCTACGAGGGAGTCGCGCTGGTGGTG 544
QY 541 ctcttgttcaactgtttgcctgaaagaaaccttgggttcataaatacggattcggc 600
Db 545 CTCTTGTCAACTTGTTTGCCCTGAAAGAAACCCCTTGGTTCATCAAAATCGGATTCGGC 604
QY 601 ctcaagatcgtctctcttgatctccttcttccatcttagccattctagcgggtcccaaccccaag 660
Db 605 CTCACGATCGCTCTCTTCGTGATCTCCTTTATGCCATTCTAGCGCGTCCAAACCCCAAG 664
QY 661 gctccaaactcacaggtatggactgcttgagagaaactatactggctgggtccgacggcgtc 720
Db 665 GCTCCAAACTCACAGGTATGGACTGCTTGGAGGCAACTATACTGGCTGGTCCGACGGCGTC 724
QY 721 tgccttcactcgtggcctttcgacatcctcgtcttcattgttcattgcttggacgcagcaatg 780
Db 725 TGCTTCATCCTGGGCCCTTTCACATCCTGCTTCATGTTTCATTGGCTTGGACGCAGCAATG 784
QY 781 catctggctgaagaatgcacagatgctgctcgtacgggtaccccaagcagtggtcagtgca 840
Db 785 CATCTGGCTGAAGAATGCACAGATGCTGCTCGTACGGTACCCAAAGCAGTGGTCAGTGCA 844
QY 841 atcataattggccttctgcacgcctttcccatatacaaatcgcagttctgtatggaattaca 900
Db 845 ATCATAAATGGCTTCTGCACCGCCTTTCCATATACAATCGCAGTTCTGTATGGAATTACA 904
QY 901 gatctcgactctattctaagtcccgccggtctattccattccagacaaatgacgcagtc 960
Db 905 GATCTCGACTCTATTCTAAGTTCGCGCGGCTATATTCATTCGAGACAATGAGGCAGTCT 964
QY 961 ctctcgtgctcagttttgcaacggtcctctcctcgtatggcgttatcgtgatggccttcttc 1020
Db 965 GCTCGAATTCGGCTTTTTCGAACGGTCTCTCTCATGTGGCGGTATCGTGATGGCCTTCTTC 1024
QY 1021 gccctcaacgctgtacaagagactcgtctcgtactcactcgtgagcgtttgcccgggacaat 1080
Db 1025 GCCCTCAACGCTGTACAAGAGACTCGTCTCGACTCACCTGGAGCTTTGCCCGGGACAAT 1084
QY 1081 gggctgggtattttccactcatctcgaacgcattcatcccccgctggcaagttcctgttgg 1140

Db 1085 GGGCTGGTATTTTCCACTCATCTCGAACGCATTCATCCCGCTGGCAAGTTCCTGTTGG 1144
QY 1141 tctctattcgcgacctggggaattcttgcccaatcgggatgtatatcttctaggttctagc 1200
Db 1145 TCTCTATTGCGGACCTGGGGAATCTTGCCACATGCGGATGTATATTCTAGGTTCTAGC 1204
QY 1201 acagcttcaaatgccttggccaattcccgccgttgaactccagcaactccttctcctgac 1260
Db 1205 ACAGCTTTCAATGCCTTGSTCAATTCCGCCGTTGTACTCCAGCAACTCTCCTTCCTGATC 1264
QY 1261 ccaatcgccctactcctctacacaaaggcagatccaaagtcttcccgagcactcgtgct 1320
Db 1265 CCAATCGCCCTACTCCTCTACCAAAAGCGAGATCCAAAGTTCTTGGCGAGCACTCGTGCT 1324
QY 1321 ttgtgttaccgcgtgggaatcggtttctctggtcaatgtgctagcgggtggttcttcacgtcc 1380
Db 1325 TTTGTGTACCGCGTGAATCGGGTTTCTGGTCAATGTGCTAGCGGTGGTCTTTCACGTCC 1384
QY 1381 gtcaccactgtgttttccagcttcccactgacgtgcctacggcgcggtcaaccatgaat 1440
Db 1385 GTCACCACTGTGTTTTTCAGCTTCCACTGACCGTGCCTACGGCCGCGTCAACCATGAAT 1444
QY 1441 tacacaagtgcgattatagcggttgacattgcttgggtgcttcttgaactgggtcgtgcat 1500
Db 1445 TACACAAAGTGCATTATAGCGCTTGACACTTGCCTTGGTGTCTTGAAGTGGGTCGTGCAT 1504
QY 1501 gccaggaagcattatcagggaacccccacttggagcttgacggacgggttcgtcgagcagaa 1560
Db 1505 GCCAGGAAGCATTATCAGGACCCCACTTGGAGCTTGACGGACGGGTCGTGGAGCAGAA 1564
QY 1561 ttccaagltgggccaatga 1578
Db 1565 TTTCAAAGTGGGCCATGA 1582

RESULT 3
SCYGL077C/c
LOCUS SCYGL077C 2475 bp DNA linear PLN 11-AUG-1997
DEFINITION S.cerevisiae chromosome VII reading frame ORF YGL077c.
ACCESSION Z72599 Y13135
VERSION Z72599.1 GI:1322592
KEYWORDS baker's yeast.
SOURCE Saccharomyces cerevisiae
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
REFERENCE 1 (bases 1 to 2475)
AUTHORS Rieger,M., Mueller-Auer,S., Brueckner,M. and Schaefer,M.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2475)
AUTHORS MIPS.
TITLE Direct Submission
JOURNAL Submitted (14-MAY-1996) Data collected by MIPS on behalf of the European yeast chromosome VII sequencing project. MIPS at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152 Martinsried, FRG; E-mail: Mewes@mips.emblnet.org

FEATURES
source
1. .2475
/organism="Saccharomyces cerevisiae"
/db_xref="taxon:4932"
/chromosome="VII"
complement(366..2057)
/gene="HNMI"
complement(366..2057)
/gene="HNMI"
/note="ORF YGL077c"
/codon_start=1
/protein_id="CAA96782.1"
/db_xref="GI:1322593"
/db_xref="SWISS-PROT:P19807"
/translation="MSIRNDNASGGYMQPDQSSNASMHKRDLRVEEIKPLDDMDSKG
AVAADGEVHLRKSFLWSILGVGFLTNSWFGISTSMVAGISSGGPMMIVYGIILVAL

ISICIGTSLGELSSAYPHAGGQFWWSLKLADPKYKRFAAYMCGSFAYAGSVFTSASTT
LSVATEVWGMVYALTHPEFIPKRWHIFVCPELLHLFLFMFFNCYKSLPIISSSSLYISL
LSFFTITITVLACSHGKFNDAKVFATFNNETGWKNNGGIAFIVGLINPAWSFSCLDCA
THMAFEVEKPERVIPAIMGTVAIGVTSFCYVVIAMFFSIODLDVLSSTTGAPILDI
YNQALGNKSGAIFLGLCLIFTSFGCVIACHTQWARLCSWAFRDNGLPLSRLWSQVNP
TGVPNAHLMSCAWITLIGLLYLASSTAFOSLITGCI AFLILSYIIPVICLLAKRNI
AHGPFWLKGFEGFFSNIVLLGWTVFSVVFSEFPVLPVTKDNMNVYCVVIGYTAISIL
YWKYKGKKEFHAESENEQAESNNFDIEDSREFSVAASDVELENEHVPWGRK"
BASE COUNT 871 a 481 c 444 g 679 t
ORIGIN

Query Match 4.2%; Score 66.4; DB 8; Length 2475;
Best Local Similarity 43.3%; Pred. No. 5.5e-08;
Matches 470; Conservative 0; Mismatches 606; Indels 10; Gaps 3;

Qy 212 cgggggggccccgactctctttacggcatcctaactcagttactctcgtctacatctgca 271
Db 1798 CTGGTGGACCCATGATGATTGTTACGGTATAAATTATTGTTGCCCTGATTTCTATTGCA 1739
Qy 272 tgcctttctcattagccgaactgacccagcgctctacccgactgccggtggccaatatcatt 331
Db 1738 TTGGTACCTCGTTGGGTGAATTATCTCCGCATATCCGCACCGCGGTGGTCACTTTTGGT 1679
Qy 332 ttgcgtcgatccctggcaccacaaaatacaatcaatcgagcatttcatacgtgtgcggactcg 391
Db 1678 GGTCTTTGAAGCTTGCCCCCTCCTAAGTACAAAAGATTTTGGCGCCTACATGTGGCGTTTCA 1619
Qy 392 tgcgtgttcttcattggtatcgctatcggaagctcagtgaccatgataccctgcacaacaga 451
Db 1618 TTGCTTACGCAGGATCCGTGTTACAAAGTGCTTCAACTACGTTATCCGTTGCTACCGAAG 1559
Qy 452 tcccggcgctgataccgccttatagtcacacatactccacaggattcgtggcaatgtctctcc 511
Db 1558 TGGTTGGTATGTACGCTCTGACGCACCCCTGAATTTCATCCCAAGAGATGGCATATATTCG 1499
Qy 512 tcatctacagaggagtcgcgctggtggtgctctctgtttcaact-----tgttgcccctgaa 566
Db 1498 TTGCTTTGAATTGTTGCATTTGTTCTTGATGTTTTTCAACTGTTTACGGTAATCCTTAC 1439
Qy 567 aagaaaccccttgggttcatgaaatcggtatccgacctcagatcgctctctcgtgatctc 626
Db 1438 CTATCATTTTCATCTCCTCTCTATACATTTCCCTACTATCCTTTTTCACAAATTACAATTA 1379
Qy 627 ctttatcgccattctagcgggtccaaaccccccaaggctccaaactcacaggatggactgc 686
Db 1378 CTGTATTGGCATGTTCTCATGGA--AAGTTCACGATGCAAAAGTTTGTGTTTGGCCACATT 1321
Qy 687 ttggagcaactatactggctggtccgacggcgtctgcttcactcctgggccccttcgacatc 746
Db 1320 TAATAAATGAACAGGTTGGAAGAACGCGGGTATCGCGTTATTGTGCGTTTGATTAACCC 1261
Qy 747 ctgcttcattgttcattggttggaagcagcaatgcattcgttggaagaatgcacagatgc 806
Db 1260 AGCTTGGTCAATTTTCGTGCCCTTGACTGTGCAACCCCATATGGCGTTTGAAGTTGAAAAACC 1201
Qy 807 tgctcgtaacgtaccccaagcagtggtcagtgcaatcaataattggtctctcgaaccgctt 866
Db 1200 AGAAGAGTTATTCCCATTTGCTATCATGGGAACAGTCCCCATTGGGTTTGTCACTTCCTT 1141
Qy 867 tccatatacaatcgcagttctgtatgggaattacagatctcgactctattcgaattccgc 926
Db 1140 TTGTTATGTTATCCCTATGTTCTTCTATACAAGATCTGGACGCTGTCTTCTTCTAC 1081
Qy 927 cggctatatcca---ttcgagacaatgacgcagtcacctcgtgcgtcagtttttgcaac 983
Db 1080 AACAGCGCCCCCAATCTTGGACATTATAATCAGGCATTTGGGTAATAAATCAGGTGGCAT 1021
Qy 984 ggtccctctcatgtggcgggtatcgatggccttcttcgcccctcaacgctgtacaagagac 1043
Db 1020 TTTCTTGGGTTGCTTGATCTCTATTACCTCTTTTGGTTGCGTCATTGCTTGTCACTTG 961
Qy 1044 tgcgtctcgaactcaactggagctttgccccgggacaatgggctgggtattttccactcatct 1103

Db 960 GCAGGCAAGGTTATGTTGGTCATTTGCCAGGACAATGGTCTTCCATTATCCCGTTATG 901
Qy 1104 cgaacgcattcatcccgctggcgaagttcctggtttggtctctattcgcgacctggggaat 1163
Db 900 GTCGCAAGTTAAACCCACACACTGGTGACCTTTGAACGCTCATTTAATGTATCGCGCTG 841
Qy 1164 tctggccacatgcgatgtatatcttctaggttcttagcacagcctttcaatgccttggtcaa 1223
Db 840 GATAACCCCTCATTTGGCCTACTTTATTGCTTCCAGTACGGCTTTTCAGTCCCTTAATTAC 781
Qy 1224 ttccgcccgttgctactccagcaactctccttctctgatacccaatgccttactcttacc 1283
Db 780 AGGTTGATTGCATTTTATTGTTATCTCTACATCATTTCCGGTTATATGTTACTTGCTAA 721
Qy 1284 aaagcg 1289
Db 720 AAAGCG 715

RESULT 4
YSCCTR YSCCTR 2725 bp DNA linear PLN 27-APR-1993
LOCUS Saccharocycles cerevisiae choline transport protein gene, complete
DEFINITION cds.
ACCESSION J05603
VERSION J05603.1 GI:171329
KEYWORDS choline transport protein.
SOURCE S.cerevisiae DNA.
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
REFERENCE 1 (bases 1 to 2725)
AUTHORS Nikawa,J.-I., Hosaka,K., Tsukagoshi,Y. and Yamashita,S.
TITLE Primary structure of the yeast choline transport gene and
regulation of its expression
JOURNAL J. Biol. Chem. 265, 15996-16003 (1990)
MEDLINE 90368823
COMMENT Draft entry and computer-readable sequence for [J. Biol. Chem.
(1990) in press] kindly provided
by J.-I.Nikawa 30-JUL-1990.

FEATURES
Location/Qualifiers
1..2725
/organism="Saccharomyces cerevisiae"
/db_xref="taxon:4932"
446..>2211
/gene="CTR"
/product="choline transport protein"
446..2392
/gene="CTR"
455..2146
/gene="CTR"
/codon_start=1
/product="choline transport protein"
/protein_id="AAA34537.1"
/db_xref="GI:171330"
/translation="MSIRNDNASGGYMQPDQSSNASMHKRDRLRVEEIKPLDDMSKG
AAADGEVHLRKSPSLWSILGVGFLNSWFGISTSMVAGISSGGPMIIVYGIIVL
ISICIGTSLGELSSAYPHAGGQFWWSLKLADPKYKRFAAYMCGSFAYAGSVFTSASTT
LSVATEVWGMVYALTHPEFIPKRWHIFVCPELLHLFLFMFFNCYKSLPIISSSSLYISL
LSFFTITITVLACSHGKFNDAKVFATFNNETGWKNNGGIAFIVGLINPAWSFSCLDCA
THMAFEVEKPERVIPAIMGTVAIGVTSFCYVVIAMFFSIODLDVLSSTTGAPILDI
YNQALGNKSGAIFLGLCLIFTSFGCVIACHTQWARLCSWAFRDNGLPLSRLWSQVNP
TGVPNAHLMSCAWITLIGLLYLASSTAFOSLITGCI AFLILSYIIPVICLLAKRNI
AHGPFWLKGFEGFFSNIVLLGWTVFSVVFSEFPVLPVTKDNMNVYCVVIGYTAISIL
YWKYKGKKEFHAESENEQAESNNFDIEDSREFSVAASDVELENEHVPWGRK"

polyA_signal 2206..2211
/gene="CTR"
polyA_signal 2387..2392
/gene="CTR"
BASE COUNT 806 a 469 c 545 g 905 t
ORIGIN

Db 17114 KGMCMYCWGRRGVCSCMTSRSAMCCSYCYAKCKSMCYCYGMSMKGYIYACSYRG 17173
QY 634 gccattctagcgcggtccaaccccaaggctccaaactcacaggtat--ggactgcttg 690
Db 17174 SMSKYCMRGSTYSTSCGCCCTTTTCCCCCNANTGGGAAGCTTTNCNKKTYSYRKR 17233
QY 691 agaacatactgctggtccgacgcgctgctgcttcctcctcctgctgcttcgacatcctgc 750
Db 17234 NGCAMCKYNNYNNWSRSCRAGMSCTKYKSSMTMSMASYCWMSMYCYCSMRSMASRGMS 17293
QY 751 ttcattgttcattgcttgacgcagcaatgcattgctgctgaagaatgcacagatgctgct 810
Db 17294 WSSYMKMKSSWRMSYMKCCWCMKCMYCNMRSMSRSGMSYMYASWKSSSRGICYCTRCY 17353
QY 811 cgtacggtacccaaagcagtggtcagtgcaatcataatgctgcttcctgcacgccttcca 870
Db 17354 WCMSSKSKCYKSYMMMRSMKRRWCKSAM--YRMRMMWKGKRGAMMMCKCYSRMSR 17413
QY 871 tatacaatcgcagttctgtatggaattacagatctcgcactctattcttaagtccgcgcgc 930
Db 17414 CAGCMKMGYMYRMRCWMSKRRWCKSAM--YRMRMMWKGKRGAMMMCKCYSRMSR 17468
QY 931 tatattccattcagacaatgacgcagtcagtccttcgctgcctcagtttgcacagcgtctc 990
Db 17469 CMMMKSYCASCRCSCAMSGTYMKCASYMCMSSYWCCTCWCCTSYNSYRCTCWCCKGWSYT 17528
QY 991 tcatgtgcggtatcgtgatgccttccttcctcctcctcctcctcctcctcctcctcct 1050
Db 17529 KSKSWSSSMSSYKGRKRSYSMCCCTSRGMSWRCWRRCCYMRGASSMRAGSMRRRAKGRSW 17588
QY 1051 cgactcacctgagagctttgcccgggacaaatggcgtggtattttccactcatctcgaacgc 1110
Db 17589 GGRSKWMTGGMWRSKYIYCTGRRMMMTYMCWCWRRRRSYMYRMSAMGMKSSWSGMR 17648
QY 1111 attaatccccgctggcaagttcctgttctgtctctattcgcgacctggggaattcggcc 1170
Db 17649 MMGSASSRRCKSASRSSWCSRRMKMGMRGSCWSSKMWGGSRRSRSSACKCKGSRGMRRSKR 17708
QY 1171 acatgcggatgtatatt 1187
Db 17709 SSKYKRGGRGKRKRSMTK 17725

RESULT 8
SPAC1039
LOCUS SPAC1039 27893 bp DNA linear PLN 15-DEC-1999
DEFINITION S.pombe chromosome I cosmid c1039.
ACCESSION AL133521
VERSION AL133521.1 GI:6594259
KEYWORDS 2,2-dialkylglycine decarboxylase; amino-acid permease;
Aminotransferases class-III; Bacterial transferase hexapeptide;
C2H2 type; esterase; family 31 glucosidase; Glycosyl hydrolases
family 31; isp5; pyridoxal-phosphate; serine acetyltransferase;
sexual differentiation process putative amino-acid permease isp5;
sugar tr; translation initiation inhibitor; transporter; UPF0076;
zf-C2H2; zinc finger protein.
fission yeast.
SOURCE Schizosaccharomyces pombe
ORGANISM Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
Schizosaccharomycetales; Schizosaccharomycetaceae;
Schizosaccharomyces.
1 (bases 1 to 27893)
REFERENCE Hupl,C., Aves,S., McDougall,R.C., Rajandream,M.A. and Barrell,B.G.
AUTHORS Direct Submission
TITLE Submitted (13-DEC-1999) European Schizosaccharomyces genome
JOURNAL sequencing project, Sanger Centre, The Wellcome Trust Genome
Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk
and Department of Biological Sciences, University of Exeter, Perry
Road, Exeter EX4 4QG, United Kingdom
COMMENT Notes:
Details of yeast sequencing at the Sanger Centre are available on
the World Wide Web.

(URL, http://www.sanger.ac.uk/Projects/S_pombe/)
During 1995 to 1996 about 66% of S. pombe chromosome 1 was
sequenced by the Sanger Centre. The sequencing of the S. pombe
genome is now being continued with funding from The European
Commission. Fourteen European sequencing laboratories, including
the Sanger Centre, are participating in the project.
Protein coding regions (CDS) have been predicted with the help of
computer analysis using the Genefinder program in PomBase (an ACEDB
database) with additional predictions for the branch-acceptor sites
supplied by the program Sp3splice. CAUTION: It is possible that for
any individual CDS we may have underestimated or overestimated the
number of introns/exons or we may not have chosen the correct
splice donor/acceptor sites.
CDS are numbered using the following system eg SPBC25H2.01c. SP (S.
pombe), B (chromosome 2), c25H2 (cosmid name), .01 (first CDS), C
(complementary strand).
The more significant matches with motifs in the PROSITE database
are also included but some of these may be fortuitous.
The length in codons is given for each CDS.
IMPORTANT: This sequence MAY NOT be the entire insert of the
sequenced clone. It may be shorter because we only sequence
overlapping sections once, or longer, because we arrange for a
small overlap between neighbouring submissions.
Cosmid c1039 is overlapped at the 5' end by cosmid c29B12, EMBL
entry SPAC29B12, accession number Z99164 and at the 3' end by
cosmid c922.

FEATURES
source
1. .27893
/organism="Schizosaccharomyces pombe"
/strain="972h-"
/db_xref="taxon:4896"
/chromosome="I"
/map="IR"
/clone="cosmid c1039"
1. .49
/note="nominal overlap with cosmid SPAC29B12 S. pombe
chromosome 1"
1375. .3078
/gene="SPAC1039.01"
1375. .3078
/gene="SPAC1039.01"
/note="SPAC1039.01, len:567,
SIMILARITY:Schizosaccharomyces pombe, O74537, putative
amino-acid permease., (557 aa), fasta scores: opt: 2304,
E():0, (62.5% identity in 520 aa)"
/codon_start=1
/label="SPAC1039.01
/product="amino-acid permease"
/protein_id="CAB63537.1"
/db_xref="GI:6594260"
/translation="MSSINEKTDVSSKHSFTFEPSSDEVKAAEIVDTHENFYDDQG
DIDNPEELAEIKYKQEFRRQLSLFGIFISFSLGMLPSVASTLVFGLWYGVPLGLW
AWLIAMFFLICVSMMAEICSAMPTSGGLYAAAVFAPKMGPLASWITGWSNYIGNI
IGQPSVNSSAASMLGAVTVNRPDVIQRWQWFLAVIQCFNCVLACLPTRTISRIN
GVATYLNATFLIAGTITILAYGGKHNHFKGDKWGYINTQWPTGFAILLSENSPI
WTMSGYDAPFHLSEECNSASVNAKPAIVMTAVIGGVGWTMGIIVAYTLTDIDSVMT
SGSMWTAYLVQAMPKALGILSLTIISAIIMQSSALIASRIAYSARDGILPFGSW
IGTVNPTYQTTPVNAVICNCIISILILFTFAGTVTLDAVFSVGAVAAFIATVPIAIR
VFETKDAFRRGPNWNLGKFSRPIGLLAVSFVALMIPILCPSPVKNPATAQEMNWTCLVY
GGPMLFTLVWVAISARKWFKPKASAHYKRPGESSDIVEGVQADIPSSSDQLKIKGD
L"
1543. .2964
/gene="SPAC1039.01"
/note="Pfam match to entry PF00324 aa_permeases, Amino
acid permease"
5742. .7547
/gene="SPAC1039.02"
5742. .7547
/gene="SPAC1039.02"
/note="SPAC1039.02, len:601,
SIMILARITY:Schizosaccharomyces pombe, O13781, hypothetical
71.4 kd protein c17g6.03 in chromosome I precursor., (635
aa), fasta scores: opt: 936, E():0, (37.7% identity in 605
aa)

```
aa)
/codon_start=1
/label=SPAC1039.02
/product="hypothetical protein"
/protein_id="CAB63538.1"
/db_xref="GI:6594261"
/translation="MKLSLP SGLGLASLLGLISSATAYSATDLTMYDWNQIIKPLE
WGQLNFHTTDTHGWLGLHLDARYKADFGEPKSFALRMKELADPKGVDLLMVDTGDL
HDGNGLSADSPQGIYTNIFTYLPYDILTIGNHELYQAAISNNTHFYFVPHWNGTYL
ASNVQIYNSSNELEQEGESTYFIFTKHGYRTLAMGFLNFESSNANNVTVPVETAIKS
EYVQQINRVDVDFLILGHIPVRDYDENKSUHASIRKVHFNTPIQILGGHSHIRDEA
VYDEASVLEGGRYCETVGLWSIDGLSANATRQYVGRPVNETRQSPNLPKIPATPL
YYTRYIDFNQRNFRHTQSEDSFDTPEGVELSKVIKQYRDDNLNLSYVFGCIPKNYY
MTEVSPAEDSIFKLMVDHILPEVLNENRSSVPHIIISNGGGVGRGSMYEGTFGPDEM
FQLNPELTNNYNYIPDPYKYAKKLYSILNGGSTRNVNDYLAALNPGYVTSDDFGED
GDDRVHTYVSTYAVPNVLQAQVGFNTTSAPETVDVVFNLNYFQTKVYLKALNTMVNETIY
TLSNVTYVWREDGKDSPPMFAQYVQQEWSDYCD"
8754..9779
/gene="SPAC1039.03"
8754..9779
/gene="SPAC1039.03"
/translation="MSQTTTETIIPLDSSVKDKLDPEYVNFYKVCNSLNPVTKTHY
PVDLRNNGNVMPGQSELPVESTEDITIPRKHTKAPSGVPSFRPHGTAPEGWPC
FVNFHGGGVWLGNIINTENSFATHMCEQAKCVVNVNDYRLAPEDFPACIDDGWEALLY
CYENATLGINPNKIAVGGSGAGNIAAVLSHKVASPANFPLVQLLVVPCDNTA
NAKTHKSWELFENTPOLPAKMMWVRRHYLPNEKDWSPSPPEYVPSSEFKNVCPAL
ICAAQCDVLSSEAIAYNEKLTKAGVESTIKIYEGCPHPVMAMDVALEKGRILNKDATN
ALLAFA"
10926..12449
/gene="SPAC1039.04"
10926..12449
/gene="SPAC1039.04"
/translation="SPAC1039.04, len:507,
SIMILARITY:Schizosaccharomyces pombe, YAOI_SCHPO, putative
transporter cld3.18c., (498 aa), fasta scores: opt: 853,
E():0, (30.7% identity in 512 aa)"
/codon_start=1
/label=SPAC1039.04
/product="putative transporter"
/protein_id="CAB63540.1"
/db_xref="GI:6594263"
/translation="MEKSISSISKASMSDEKLDKKEMDQQLPIDFGEDEVTTVEYVI
LDHKAERRLCRKFDRIPLIALLYLNFALDKSNVSNKTNMGDKDLGFVGDQYNIMI
STIPIPVLCAPFYSIYKRFGAARILPFLMSFGAMSLCQAAVKNFNGMMVAVRWFLG
MAESAVLPGVVYLYTTEYRTELARRLAIFYAANVSSAFGGLLAYGVFHIKGGKLOG
WOYLFLEGGVTFCAIVIFLVLPSVETANFLTDEEKTAKMRINDSSSAISEKLS
FKQSLTVEKHPIALILWLEEMALGVPLNSINWLPQIIVAAAGFSSVNTNMTVAPATS
GAILLVFAFISDFLNKRGIVLIAISTWIGFIVYGSIDIMNHIGVSFYFACFLMTAG
AAASSVLTSTWYNNTPENESRRAVFTSVGVPLANVMGLVSNIFRPQDAPKYVPALGI
TAGFGGLGILLVASISVYMFDFNRRRDNAQGVKKTFAADVSTKDLGEGPANPNRWEL"
11127..12374
/gene="SPAC1039.04"
/translation="Pfam match to entry PF00083 sugar_tr, Sugar (and
other) transporter"
complement(12770..15420)
/gene="SPAC1039.05c"
complement(join(12770..12813,12865..13497,13553..13851,
13897..14018,14072..15198,15300..15420))
/gene="SPAC1039.05c"
/translation="SPAC1039.05c, len:781,
SIMILARITY:Schizosaccharomyces pombe, AAF13704, c2h2 zinc
finger protein zasla., (845 aa), fasta scores: opt: 1228,
E():0, (32.7% identity in 842 aa)"
/codon_start=1
```

```
/label=SPAC1039.05c
/product="C2H2 zinc finger protein"
/protein_id="CAB63541.1"
/db_xref="GI:6594264"
/translation="MTKAKKSRAYQEGDIRYKDFQGGTKSFTRKEHARRHFRSHSTNS
KAFICPHCSSSFTSRDVLNRHVQKHVKONDAGSQTHPKHDKTLLISSDQELPFPYSYV
EADVQDAAYTDLKSIDQIPILDQPIISDNFENEGLSAIAOTGNMYFTTSRRNNSIS
GSITIPESDQIINLONKFLSTPSIKNNANKSLSPQDSISIFGTSFNFAYQQNTDNFVC
WLFNDMEKDAPTESVSNFTDGINAKQNLNMPVLDSPDIDFLSSHFVRVDKLVAKDLLS
LSAYSSLIQVMNNYQLSEELQYITRSRVNLWIAHWKDFHPRWFLHRTGLKVDDEA
PVELLAMITGMHFVGDFAFYSIAVSIHSTLRFSTYHPDFKPPASLWVYQALLIAE
IFEKMTSTVEQHNLSQLIEHGVTFESMQNLSSKDTVTEKIPKNMTGTNIAQHKHQWV
DREASKRTAFPSFVLDSQHVILFGYRPLIDITSVGLDCLDEALWNADSYEAWTSLN
ENDPPHPPVLKMFELTNEHLPPKLSPPNMIVLHGLTTIGWILSRENLSGIVENIMONN
GTNLKNRILLKASYKFWLRTYRVFPLNNGTLPLNHPYVVRGCLATYELAHISLHTNIV
ALQTYAKSIVSTSRRLYASTSRVVSANLASDDSEVSIKHAVDMVEAFLGSDMEYDVQN
ETGLHRPWCLYVSTLILWAYGYVSDGRCEMLEPNNCKSQLNAYLMQMRTGLSEKAK
DHVVVRSKTLPLLKCVIDVLCPARWGLLVGDGVRILSKLVQL"
complement(12814..12826)
/gene="SPAC1039.05c"
/translation="CTAACAGTTTtag, splice branch and acceptor"
complement(12859..12864)
/gene="SPAC1039.05c"
/translation="Gtaagt, splice donor sequence"
complement(13498..13513)
/gene="SPAC1039.05c"
/translation="CTAACAGTgttttaag, splice branch and acceptor"

Query Match 2.8%; Score 44.8; DB 8; Length 27893;
Best Local Similarity 42.8%; Pred. No. 0.15;
Matches 281; Conservative 0; Mismatches 372; Indels 3; Gaps 1;

Qy 671 cacaggatatggactgcttggagcaactatactatggctgggtccgacggctctgtcttcaccc 730
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2123 CAAGAATTTGGGAGATTATTAACACAAACCAATGGCCTACCGGTTTGTCTATCCTTC 2182

Qy 731 tgggcttbtgcacatcctgcttcatgttctatggcttggacggcagcaatgcacatctggctg 790
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2183 TTTTCTTCAACAGTCTCTATCTGGACCATGTCTGTGTACGATGCCCTTTCATCTTAGTG 2242

Qy 791 aagaaatgcacagatgctgctgcgtacgggtacccaaagcagtggtcagtgccaatcataattg 850
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2243 AAGAATGTTCCAATGCAAGTGTAAATGCTCCAAAGGCCATTGTAATGACTGCAGTTATTG 2302

Qy 851 gcttctgcacccgcttccatatacaaatgcagcttctgtatggaaatcacagatctcgact 910
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2303 GAGCTGTAGTTGGATCGATTATGCAAAATATCTCGCTGCTATACCTTAACTGATATTGACT 2362

Qy 911 ctattctaaagtctccgctctatatctccattcgagacaaatgacgacgctcccttcggtcgc 970
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2363 CGGTTATGAACACATCTCGATCTATGTGGACGGCTTATCTTGTACAAGCAAT---GCCTC 2419

Qy 971 tcsagtttgcaacggctctctcatggtggcggtatcgatggcctcttcgacctcaacg 1030
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2420 CTAAGGCTGCCTTGGGTATATTGAGTCTTACGATATTCTTGTCCATCATCATGGTCAAA 2479

Qy 1031 ctgtacaagagactgctctcgactcacctggagcttggccgggacaaatggcggttat 1090
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2480 GTGCACTTATCGCTTCTATCTCGTATCGCTATTCCTATGCCAGAGCGGTATTCTTCCCT 2539

Qy 1091 ttctcactcatctcgacgcgacttcctcccttccttccttccttccttccttccttccttcct 1150
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2540 TTTTCAAGGTTGGATCGGACGGTCAACCCATACACTCAAACTCTCTGTAAATGCCGTAATAT 2599

Qy 1151 cgcactgggggaattctggccacatcgggatgtatatcttctaggttcttagcacagctttca 1210
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2600 GTAACATGCATCATTTAGTATCTCTTATTTTGTTCCTTACGTTTGTGTTGACAGTACACCTTAG 2659

Qy 1211 atgcttgggtcaattctcgccggttgactcactcagcaactctctctctctctctctctctctc 1270
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2660 ATGCTGTATTTTCTGTGCGGTGCTGTTGCTGCTTTTATTTGCTTCACTGTTTCCAATCGCTTA 2719

Qy 1271 tactctctaccaaaagcgagatccaaagctcttcttgcgagcactgctgtcttttctgtg 1326
```


Db 2720 TCGTGTCTCTTTACCAAGATGCTGATTTCCTAGAGGTCTTGGAATTGGGG 2775

RESULT 9

AP003687/c

LOCUS AP003687 147203 bp DNA linear PLN 25-AUG-2001

DEFINITION Oryza sativa genomic DNA, chromosome 1, PAC clone:P0660F12.

ACCESSION AP003687

VERSION AP003687.3 GI:15290149

KEYWORDS

SOURCE Oryza sativa (cultivar:Nipponbare) DNA, clone:P0660F12.

ORGANISM Oryza sativa

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.

REFERENCE 1 (bases 1 to 147203)

AUTHORS Sasaki,T., Matsumoto,T. and Yamamoto,K.

TITLE Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC clone:P0660F12

JOURNAL Published Only in Database (2001) In press

REFERENCE 2 (bases 1 to 147203)

AUTHORS Sasaki,T., Matsumoto,T. and Yamamoto,K.

TITLE Direct Submission

JOURNAL Submitted (23-MAY-2001) Takuji Sasaki, National Institute of Agrobiological Sciences, Rice Genome Research Program; Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan

(E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/, Tel:81-298-38-7441, Fax:81-298-38-7468)

COMMENT On Aug 24, 2001 this sequence version replaced gi:14625023.

Genes were predicted from the integrated results of the following: GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as SplicePredictor (October 1998 version). The genomic sequence was searched against NCBI Nonredundant protein database, nr (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at RGP. Protein homologies of the coding regions were searched against NCBI Nonredundant protein database with BLASTP2.0. ESTs represent the identified cDNA sequences using BLASTN 2.0 with the corresponding DBJ accession no. and RGP clone ID.

A gene with identity or significant homology to a protein is classified based on the protein name to indicate the homology level such as same name, 'putative-' and '-like protein'. A gene without significant homology to any protein but with EST homology (covering almost the entire length of partial sequence) is classified as an 'unknown' protein. A gene predicted with a gene prediction program is classified as a 'hypothetical' protein.

The orientation of the sequence is from SP6 to T7 of the PAC clone. Detailed information on overlap and assembly quality together with annotation of this entry is available at http://rgp.dna.affrc.go.jp/GenomeSeq.html.

FEATURES

source

1..147203

/organism="Oryza sativa"

/cultivar="Nipponbare"

/db_xref="taxon:4530"

/chromosome="1"

/clone="P0660F12"

1110..3179

/gene="P0660F12.1"

1110..3179

/gene="P0660F12.1"

/note="probably inactive due to frameshift in CDS

pseudogene

similar to Oryza sativa chromosome 10, OSUNBa0046L02.1"

/pseudoc

join(6037..6069,6882..7160,7364..7783)

/gene="P0660F12.2"

join(6037..6069,6882..7160,7364..7783)

/gene="P0660F12.2"

/note="contains Est AU102097(C10257)

unknown protein"

/codon_start=1

/protein_id="BAB63840.1"

/db_xref="GI:15290150"

/translation="MSTSLPAAEKKAARKKALKGWDPTFIKRNFFPWRSVLGAQLSV SNGAADEKSMLHRPQFRNSVSNESSSLNHNISFLGRIOOLSGTHESRRYSTAAFRML TKWHVAVDNHRSAKYAAHVFCGLKVKVNIHVTKTLFKVMLSTLNYDEMQHAYCCI RLEPSTREMTCTVPACCTCTRFNVSLNTEQVKVAKCHFSLSAISSTYIRVMILECYW ATWQPEKLAKHONFCHVSIKYKL"

join(11112..11244,11317..11524,12485..12567,12599..12729)

/gene="P0660F12.3"

join(11112..11244,11317..11524,12485..12567,12599..12729)

/gene="P0660F12.3"

/note="contains ESTs C27575(C52254),AU0823350(C52254)

unknown protein"

/codon_start=1

/protein_id="BAB63841.1"

/db_xref="GI:15290151"

/translation="MSTSEAGLPDPATAASSLVDCRLPGREGCEPLHSSSPLSSHLH GSGRGPLRQYNATLHCLPHFPFSLTSPRCQDGEKSSSLGTTKATTAEGMGYGVRR RROQGTNYGKVCCKGEANFIFQNEHSHVKLSSAFILGVQMGNEQRCQKEIEDTKH GLESGSITYDPAKLKFSDDIKDSR"

complement(14943..15441)

/gene="P0660F12.4"

complement(14943..15441)

/gene="P0660F12.4"

/note="probably inactive due to frameshift in CDS

pseudogene

similar to non-phototropic hypocotyl 3"

/pseudoc

complement(join(19029..19318,19358..19489,20047..20191,20796..21003,22733..22989))

/gene="P0660F12.5"

complement(join(19029..19318,19358..19489,20047..20191,20796..21003,22733..22989))

/gene="P0660F12.5"

/note="hypothetical protein"

/codon_start=1

/protein_id="BAB63842.1"

/db_xref="GI:15290152"

/translation="MRAGAWLPTTGDMLTNDGALGGGGSGGSRITRAGAAARR GESDVAEGSASILWTHAHSRGGMTPAVIAMVYRGAATSPQPSADGSGDRADGRS SRARVKIERRGRILGGVGRKRREERESRATHVGEEMQARWPDSQAQNMRELGDAA TMLEVDEEDPERLTMTSADDEGVARTAMKTTGNIVTHRGDTVKQVHRREKKQDRVS RVRPGGASRRQRAQVVTETKGGVGGGRGWADGVVSTVKTARQREPQDAGCGPR RGEETEDEPANRLSGSGGGGWRRLGIRRSATWTERLIGIRPSGPRRNSGDGAGETDA VASKITRST"

complement(join(23330..23697,23705..24026))

/gene="P0660F12.6"

complement(join(23330..23697,23705..24026))

/gene="P0660F12.6"

/note="hypothetical protein"

/codon_start=1

/protein_id="BAB63843.1"

/db_xref="GI:15290153"

/translation="MDWVKRHMGSHGKGARGEEAPGSDGSQRRNNAGERRQAGAGR HRKREGLKAVPSTGSGEAAARWRPCGGMGGTARTPRPARRPAVAASRGDSLGLGAR RYEGFGDRWRMRMARGAGVASGRGKRRKEGEQAGGERDDQPAVTRDYARRREQSR RERVTRRPGGRRRRRFLAPAVAKQKRCELTDERRHGRETRPRRGAEMARPGKDS TAAASAALGKG"

complement(join(24662..24693,25292..25415,25920..26116,26546..26669,27218..27346,31082..31192))

/gene="P0660F12.7"

complement(join(24662..24693,25292..25415,25920..26116,26546..26669,27218..27346,31082..31192))

/gene="P0660F12.7"

/note="hypothetical protein"

/codon_start=1

/protein_id="BAB63844.1"

/db_xref="GI:15290154"

/translation="WRNSSRSEGDISOILEMEEIYTVDESLLKKKKNGKEIQNTPMA EDGALSIVITLVMPFSSENKDEDEVFYIKQGGINVINWTKIIRHYTKYTSQSQSDTI TRFGPTCQSKPPKIESQSCSSGQPTDITDANTNNSSSGTNGLTPTNTGNTFR ELGRNLNLHFTSTSPSELAELYIYRAPWSPSPCGSNTSGDLGLSRQISKANTGLPY ANERGPPTTGSSRIFEKSA"

32084..34153

gene

QY 92 agtccctgaacgttcctctggagagaacaacatttggcaccatcacccatgctgctcttg 151
: : : ||| | | : : : : : | : : : : :
Db 51507 RYAYASTGAAMSMCGTCYMKMCTRRMMATACAWRRRAWMSCCMRKCTSKWSKMGR 51448
QY 152 cctttgatttgcacacagttgggtgtatctcaggcagttcctcagctgcctactag 211
:
Db 51447 MRYCWSWARWYYAGSYMCKRSRWGGYKGRSRSRSGMSWTRGYSYARCYMSKSAGGA 51388
QY 212 cggggggcccgctcaactccttttacggcactcctaatacagttactctctgctacatctgca 271
: : | | : : : : : : : : : : | | : : : : :
Db 51387 GKCKGAGSYAKGSAGWRWGSYKGAWYSCNMSASKKSRMTCYWSCCCTSCCKCYWYWSW 51328
QY 272 tcgctttctcattagccgaactgaccagcgctctacccgactgccggtggccaaatatcatt 331
: | : : : : : : : : : : : : : : : : : : :
Db 51327 KYGAGMCWYMRWCCTCMARAWRARMARMAAMRRARMAGAAARAWMSMMWMMMRBHAAAMM 51268
QY 332 ttgcgtcgatccctggcaccacaaatcaatcgaatcgagcatttcatacgtgtgcggaactcg 391
Db 51267 ARWAAAKCWWYYSYRKYCWNMMMRASWAAMAMMMCTAAWYKKW--TRMRWWSKYSSS 51210
QY 392 tgcgtgttcttcattgctctatcggaagctcagtgaccatgatacctgctcacaaga 451
:
Db 51209 YAYAWYATMCWYCYSSMYAARMGTYSGGKKKWAISTSCMCCRRGCCMKKKGMRWCS-- 51150
QY 452 tcccgcgctgtagcgcgcttatagtcacacatactcaccagattcgtggcattgcttcc 511
| : | :
Db 51151 TTYTGMSGGKSRWWTSSMACWYWKGSRRRTKSMATSGCMRMGAMRSKGGMRKYWSCRYK 51092
QY 512 tcatctacagggagtcgcgctgggtgctcttgttcaactgttgcctgaaagaa 571
: : : : : | : : : : : : : : : : : : : : :
Db 51091 GMRWGRWSGYKGCSGYMSMGYSGRTSKYCSYTGGYCSCSAKMKCKSKSTSKCCTKSK 51032
QY 572 acccttgggttcatagaatcggaattcggcctcagcatcgctctctctgtagtctcttta 631
: : : : : | : : : : : : : : : : : : : : :
Db 51031 YSCSMGRYSSYCTCWGGRMCKGMK~-SSCCASRSKGCKSYKMGAGSWRCRRRRGMSCWYWSS 50974
QY 632 tgcattctagcgcggtccacacccccaaaggtcccaactcagaggtatggactgcttga 691
:
Db 50973 MKSYSMASYKSSARGMKCYKTSCKMYYYSMHBKSSHDBSGYKSYRKSYYBYWTCMKNKDS 50914
QY 692 gcaactatactggctggtccgacggcgctgctcttcattccttgccttgcacatcctgct 751
| :
Db 50913 MCWSVDSHMRDMYDKMRMKVSSGTRGYHBHSAWMCKMSMSRSDSGYSGMYHMSYWK 50854
QY 752 tcatgttcatggcttggacgagcaaatgcattggtggaagaatgcacagatgctgctc 811
: : : : : | : : : : : : : : : : : : : : :
Db 50853 SASKHMYYSCKMSSVYBMDCNMYSMTSHSKSVRRRWGSGSYBRYSRYKCAVMRRYBHRSSSTR 50794
QY 812 gtacgggtaccacaaagcagtggtcagtgcataca 844
:
Db 50793 SVMYGGKRSCWVRNDSVSRRRSVHTSMSSWRMCR 50761

RESULT 11
AF053231
LOCUS AF053231 1996 bp mRNA linear PLN 28-MAR-1998
DEFINITION Neurospora crassa amino acid permease (aap2) mRNA, complete cds.
ACCESSION AF053231
VERSION AF053231.1 GI:2995850
KEYWORDS
SOURCE Neurospora crassa.
ORGANISM Neurospora crassa
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariales; Sordariaceae; Neurospora.
REFERENCE 1 (bases 1 to 1996)
AUTHORS Margolles-Clark,E. and Bowman,B.J.
TITLE Direct Submission
JOURNAL Submitted (11-MAR-1998) Biology, University of California, Santa Cruz, 444 Sinsheimer Laboratories, Santa Cruz, CA 95064, USA
FEATURES
source
1 .1996
/organism="Neurospora crassa"

/db_xref="taxon:5141"
/chromosome="LGI"
/map="close to R15.3, AP31a.3"
1. .1996
/gene="aap2"
100. .1755
/gene="aap2"
/function="amino acid transporter"
/note="similar to GABA permeases; specificity unknown"
/codon_start=1
/product="amino acid permease"
/protein_id="AAC08355.1"
/db_xref="GI:2995851"
translation="MSFPPNPKSADATIQITEMTROGTPSSGEAAASTPSTSTESGD
KALEALGYTPVFKREESRWSSFSFAVSGVGTLMSTWIYGLQAGGAAAIMMSWIIIG
GAGGALAYSIAETASAYPSSGAMYFTLKFLAPEQVPLCWIAGYDLNVGTVAGGAS
TEYAAQMLLAASVJTSNFSYVPTPTHVVGVMIGLTTTHAMINTLPTAWLNRLTSGYV
VFHISVLLGACVTLVQKRHMDHLKYAFTNFQPSGWSPPGFALFGLTPAWIMTG
CDGTARIAEEAKPMQVVPRAIANATFTYVIGFFENLVVCMGDPKDLINSPSGQP
VAQLFNGMGRAPAFFTLGCGFVNLVAIPGMQAGSKTIFALSRDNLFPFHSIWVRI
SKRSOTPLIAVWTYAVLEIIINLLGLASSSTAIGAVENVCTVALNVSVIPIICKMVG
RMOKGPHMGKYVWVNAFAVAWNTENAVIEFFPTRVPTPENMNYAIVVFEFVLILA
LVEWYTHGRHYTCGLTHSPRATDMSVTRPWSVIGTSCCTL"
BASE COUNT 423 a 539 c 477 g 557 t
ORIGIN
Query Match 2.5%; Score 40; DB 8; Length 1996;
Best Local Similarity 51.7%; Pred. No. 3;
Matches 91; Conservative 0; Mismatches 85; Indels 0; Gaps 0;
QY 1350 ggteaagtgtcgtggtggtcttcacgtccgtcaccaactgtgtttttcagcttccact 1409
||||||| | | ||||| | | | | | | | | | |
Db 1494 GGTCATGCGTTTGCCTGTCCTGGATACGTTTCATGCGCGTGATCTTCTTCTTCCCCAC 1553
QY 1410 gacgtgcctacggcgcggtcacaacctgaattacacagtgcgattataggcgttgcaact 1469
| | | | | | | | | | | | | | | | | | | |
Db 1554 GCGGTACCTGTTACACCTGAGAAATATAACTACGCCCATAGTGTCTTTTCTTCTCGTTCT 1613
QY 1470 tgccttgggtgcttgaactgggtcgctgcatgccagggaagcattatcagggaacccc 1525
| | | | | | | | | | | | | | | | | | | |
Db 1614 CATTTAGCCCTCGTTTCTGTGTACACCCATGGACGCCCACTACTATACCGGCCCCC 1669
RESULT 12
CNS01D5D 720 bp mRNA linear PLN 03-SEP-1999
LOCUS Botrytis cinerea strain T4 cDNA library under conditions of
DEFINITION nitrogen deprivation.
ACCESSION AL116393
VERSION AL116393.1 GI:5831609
KEYWORDS cDNA library; nitrogen deprivation.
SOURCE Botryotinia fuckeliana.
ORGANISM Botryotinia fuckeliana
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Leotiomycetes;
Helotiales; Sclerotiniaceae; Botryotinia.
REFERENCE 1 (bases 1 to 720)
AUTHORS Bitton,F., Levis,C., Fortini,D., Pradier,J.M. and Brygoo,Y.
TITLE Direct Submission
JOURNAL Submitted (01-SEP-1999) Phytopathologie, INRA, route de St Cyr,
78026 Versailles, France
REFERENCE 2 (bases 1 to 720)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (01-SEP-1999) Genoscope - Centre National de Sequencage :
CP 5706 91057 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT The cDNA library to be analyzed within the framework of this
project was created using a Botrytis cinerea strain which was grown
under conditions of nitrogen deprivation, which is the normal
situation for B. cinerea during its development on its host plant.
The library was produced in an oriented direction, in the pBSII
vector.

Db 5700 YSYDASTCYTSRGTBCTTBYSTBTBNCRCNCYSDSTYRANCYSSSTYDATBNSTY 5646

Search completed: April 26, 2002, 23:09:34
Job time: 13563 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2002, 02:58:44 ; Search time 4999.95 Seconds
(without alignments)
4259.682 Million cell updates/sec

Title: US-09-882-694A-7
Perfect score: 1578
Sequence: 1 atggactccagaccaaagtgg.....aatttcaagttgggcccatga 1578

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 13736207 seqs, 6748477542 residues

Word size : 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST: *
1: em_estba: *
2: em_esthum: *
3: em_estin: *
4: em_estmu: *
5: em_estov: *
6: em_estpl: *
7: em_estro: *
8: em_htc: *
9: gb_est1: *
10: gb_est2: *
11: gb_htc: *
12: gb_gss: *
13: em_gss_hum: *
14: em_gss_inv: *
15: em_gss_pln: *
16: em_gss_vrt: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query	Length	DB	ID	Description
1	22	1.4		728	10	BE272788	BE272788 601104991
2	21	1.3		832	10	BE299324	BE299324 601119739
3	20	1.3		243	9	AW408618	AW408618 UI-HF-BM0
4	20	1.3		272	10	BF933836	BF933836 CM2-NT024
5	20	1.3		277	10	BE244343	BE244343 TCBAP2E03
6	20	1.3		285	10	BM152000	BM152000 TCBAP1E11
7	20	1.3		343	9	AA644607	AA644607 af74b07.r
8	20	1.3		345	9	AA629140	AA629140 af60b06.s
9	20	1.3		363	10	BI316879	BI316879 saf77e08.
10	20	1.3		365	9	AW407577	AW407577 UI-HF-BM0
11	20	1.3		365	10	BM152194	BM152194 TCBAP1E61
12	20	1.3		391	9	AW222767	AW222767 EST299578
13	20	1.3		434	9	AL120231	AL120231 DKFp761G
14	20	1.3		455	10	BM174982	BM174982 TgSTzyb1
15	20	1.3		455	10	BE241865	BE241865 TCAAP2E00
16	20	1.3		466	10	BG395808	BG395808 602458985
17	20	1.3		469	10	BM153121	BM153121 TCBAP1Q13

18	20	1.3	471	10	BF043417	BF043417 BP2500068
19	20	1.3	482	10	BM049825	BM049825 603624169
20	20	1.3	490	10	BI760416	BI760416 603045426
21	20	1.3	538	9	AW077262	AW077262 fj33g10.y
22	20	1.3	583	10	BG478113	BG478113 602522764
23	20	1.3	586	10	BE900139	BE900139 601673051
24	20	1.3	591	10	BM490083	BM490083 ppg2n.pk0
25	20	1.3	606	10	BE384337	BE384337 601277428
26	20	1.3	620	10	BE384039	BE384039 601272830
27	20	1.3	630	10	BG761867	BG761867 602718074
28	20	1.3	634	10	BE900587	BE900587 601673473
29	20	1.3	644	10	BG720272	BG720272 602692334
30	20	1.3	648	10	BI772780	BI772780 603058910
31	20	1.3	656	10	BI391871	BI391871 ppgln.pk0
32	20	1.3	662	10	BG717889	BG717889 602693931
33	20	1.3	667	10	BF204617	BF204617 601867037
34	20	1.3	675	10	BF025655	BF025655 601670506
35	20	1.3	679	10	BE887718	BE887718 601511882
36	20	1.3	682	10	BI765819	BI765819 603047623
37	20	1.3	683	10	BG724330	BG724330 602697983
38	20	1.3	688	10	BE294290	BE294290 601172983
39	20	1.3	690	10	BI753384	BI753384 603026565
40	20	1.3	724	10	BE887301	BE887301 601510125
41	20	1.3	729	10	BG257589	BG257589 602377252
42	20	1.3	733	10	BE281295	BE281295 601155982
43	20	1.3	740	10	BF204036	BF204036 601869161
44	20	1.3	744	10	BI459520	BI459520 603200525
45	20	1.3	752	10	BG819515	BG819515 602783340

ALIGNMENTS

RESULT 1
BE272788 728 bp mRNA linear EST 13-JUL-2000
LOCUS 601104991f1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:3347735 5',
DEFINITION mRNA sequence.
ACCESSION BE272788
VERSION BE272788.1 GI:9147139
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 728)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DMP
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LICM139 row: f column: 24
High quality sequence stop: 646.
Location/Qualifiers
1. 728
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3347735"
/clone_lib="NIH_MGC_14"
/tissue_type="renal cell adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: kidney; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of

California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 126 a 229 c 248 g 125 t

ORIGIN

Query Match 1.4%; Score 22; DB 10; Length 728;
Best Local Similarity 100.0%; Pred. No. 3.6;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 503 atgtcttcctcatctacgagg 524
|||||

Db 164 ATGTCTTCCTCATCTACGAGG 185

RESULT 2

BE299324

LOCUS 60119739F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:3029577 5',
EST.

DEFINITION mRNA sequence.

ACCESSION BE299324

VERSION BE299324.1 GI:9183072

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC

CDNA Library Preparation: Ling Hong/Rubin Laboratory

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov

Plate: LLCM89 row: n column: 10

High quality sequence start: 11

High quality sequence stop: 708.

FEATURES Location/Qualifiers

1..832

organism="Homo sapiens"

db_xref="taxon:9606"

clone="IMAGE:3029577"

clone_lib="NIH_MGC_17"

tissue_type="rhabdomyosarcoma"

lab_host="DH10B (phage-resistant)"

note="Organ: muscle; Vector: pOTB7; Site_1: EcoRI; Site_2: XhoI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 148 a 232 c 276 g 176 t

ORIGIN

Query Match 1.3%; Score 21; DB 10; Length 832;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 503 atgtcttcctcatctacgagg 523
|||||

Db 198 ATGTCTTCCTCATCTACGAGG 218

RESULT 3

AW408618

LOCUS 243 bp mRNA linear EST 16-FEB-2000

DEFINITION UT-HF-BM0-adv-d-06-0-UI.r1 NIH_MGC_38 Homo sapiens cDNA clone IMAGE:3062843 5', mRNA sequence.

ACCESSION AW408618

VERSION AW408618.1 GI:6927675

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Eco RI site shown at the beginning of the sequence.
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
CDNA Library Preparation: M.B. Soares Lab
CDNA Library Arrayed by: M.B. Soares Lab
DNA Sequencing by: M.B. Soares Lab
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

Seq primer: M13 Forward.

Location/Qualifiers

1..243

organism="Homo sapiens"

db_xref="taxon:9606"

clone="IMAGE:3062843"

clone_lib="NIH_MGC_38"

tissue_type="lymph"

cell_type="germinal center B cells"

cell_line="MGC85"

lab_host="DH10B (LTI)"

note="Vector: pT7T3-Pac; Site_1: NotI; Site_2: Eco RI; Constructed from size fractionated cytoplasmic mRNA (2.5-3.5kb). Directionally cloned. Cells provided by Louis M. Staudt, Ph.D. Library preparation by Maria de Fatima Bonaldo, Ph.D. and M. Bento Soares, Ph.D."

BASE COUNT 35 a 59 c 102 g 37 t

ORIGIN

Query Match 1.3%; Score 20; DB 9; Length 243;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 505 gtcttcctcatctacgagg 524
|||||

Db 200 GTCTTCCTCATCTACGAGG 219

RESULT 4

BF933836/c

LOCUS CM2-NT0244-211200-642-e02 NT0244 Homo sapiens cDNA, mRNA sequence.

DEFINITION BF933836

ACCESSION BF933836.1 GI:12351160

VERSION BF933836.1

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.

AUTHORS 1 (bases 1 to 272)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL
MEDLINE
COMMENT

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM2&tl2=CM2-NT0244-
211200-642-e02&tl3=2000-12-21&tl4=1)
Seq primer: puc 18 forward
High quality sequence start: 22
High quality sequence stop: 272.

FEATURES

source

1. .272
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NT0244"
/dev_stage="Adult"
/note="Organ: nervous_tumor; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
38 a 70 c 79 g 85 t

BASE COUNT
ORIGIN

Query Match 1.3%; Score 20; DB 10; Length 272;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 505 gtcttctctcatctacgagg 524
|||||
Db 251 GTCTTCTCATCTACGAGG 232

RESULT 5

BE244343
LOCUS
DEFINITION
TCBAP2E0330 Pediatric pre-B cell acute lymphoblastic leukemia
Baylor-HGSC project-TCBA Homo sapiens cDNA clone TCBAP0330, mRNA
sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
human.
ORGANISM
Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 277)
Wei, Y., Tsang, Y.T.M., Mei, G., Ku, J.M., Ali-Osman Jr., F.R., Muzny, D.,
Bouck, J., Gibbs, R.A. and Margolin, J.F.
Pediatric Leukemia cDNA Sequencing Project
Unpublished (2000)

Contact: Dr. Judith F. Margolin
Texas Children's Cancer Center and Human Genome Sequencing Center
at Baylor College of Medicine
1102 Bates, MC3-3320 Houston, TX 77030, USA
Tel: 832-824-4536
Fax: 832-825-4038
Email: clones@txccc.org
Citation: Carninci, P. and Hayashizaki, Y. High efficiency
full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Seq primer: M13 primer.
Location/Qualifiers
1. .277

FEATURES

source

1. .272
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NT0244"
/dev_stage="Adult"
/note="Organ: nervous_tumor; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
38 a 70 c 79 g 85 t

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="TCBAP0330"
/clone_lib="Pediatric pre-B cell acute lymphoblastic
leukemia Baylor-HGSC project-TCBA"
/sex="male"
/tissue_type="leukopheresis"
/cell_type="pre-B cell"
/dev_stage="pediatric 2 years"
/lab_host="DH10B"
/note="Vector: lambda pSB; Site_1: BamHI; Site_2: EcoRI;
First strand cDNA was primed with an anchored
XhoI-oligo(dT) primer [5'GGAGGACTCGAGCGGCCGAGGAG(T)VN
3'; V=A,C,G; N=A,C,G,T] and then dG tailed. Second strand
was primed with a BamHI-dC primer
[5'AGAGAGCTCGGATCCGCGCGCAATAATAAT(C) 3'].
Double-stranded cDNA was then digested with BamHI and XhoI
and directionally cloned into the BamHI and SalI sites of
lambda pSB vector. Library went through one round of
normalization. Library was constructed by Wei Yu at RIKEN
of Japan (Carninci P, Westover A, Nishiyama Y, Ohsumi T,
Itoh M, Nagaoka S, Sasaki, Okazaki Y, Muramatsu M,
Schneider C, Hayashizaki Y, High efficiency selection of
full-length cDNA by improved biotinylated cap trapper.,
DNA Res 4: 1, 61-6, Feb 28, 1997)"
47 a 79 c 97 g 54 t

Query Match 1.3%; Score 20; DB 10; Length 277;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 505 gtcttctctcatctacgagg 524
|||||
Db 159 GTCTTCTCATCTACGAGG 178

RESULT 6

BM152000
LOCUS
DEFINITION
BM152000 285 bp mRNA linear EST 30-NOV-2001
TCBAP1F11845 Pediatric pre-B cell acute lymphoblastic leukemia
Baylor-HGSC project-TCBA Homo sapiens cDNA clone TCBAP1184, mRNA
sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
human.
ORGANISM
Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 285)
Wei, Y., Tsang, Y.T.M., Mei, G., Ku, J.M., Ali-Osman, F.R. Jr.,
Gunnaratne, P.H., Muzny, D., Bouck, J., Gibbs, R.A. and Margolin, J.F.
Pediatric Leukemia cDNA Sequencing Project (2001)
Unpublished (2001)
Contact: Dr. Judith F. Margolin
Texas Children's Cancer Center and Human Genome Sequencing Center
at Baylor College of Medicine
1102 Bates, MC3-3320 Houston, TX 77030, USA
Tel: 832-824-4536
Fax: 832-825-4038
Email: clones@txccc.org
Seq primer: M13 primer.
Location/Qualifiers
1. .285

FEATURES

source

1. .285
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="TCBAP1184"
/clone_lib="Pediatric pre-B cell acute lymphoblastic
leukemia Baylor-HGSC project-TCBA"
/sex="male"
/tissue_type="leukopheresis"

/cell_type="pre-β cell"
/dev_stage="pediatric 2 years"
/lab_host="DH10B"
/note="Vector: lambda pSB; Site_1: BamHI; Site_2: EcoRI;
First strand cDNA was primed with an anchored
XhoI-oligo(dT) primer [5'GGAGGACTCGAGCGCGCCGAGGAG(T)VN
3'; V-A,C,G; N-A,C,G,T] and then dG tailed. Second strand
was primed with a BamHI-dC primer
[5'AGAGAGCTCGATCCGCGCGCGCAATAATAATAT(C) 3'].
Double-stranded cDNA was then digested with BamHI and XhoI
and directionally cloned into the BamHI and SalI sites of
lambda pSB vector. Library was constructed by Wei Yu at RIKEN
normalization. Library was constructed by Wei Yu at RIKEN
of Japan (Carninci P, Westover A, Nishiyama Y, Ohsumi T,
Itoh M, Nagaoka S, Sasaki, Okazaki Y, Muramatsu M,
Schneider C, Hayashizaki Y, High efficiency selection of
full-length cDNA by improved biotinylated cap trapper.,
DNA Res 4: 1, 61-6, Feb 28, 1997)"
BASE COUNT 47 a 73 c 112 g 53 t
ORIGIN

Query Match 1.3%; Score 20; DB 10; Length 285;
Best Local Similarity 100.0%; Pred. NO. 32;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 505 gtcttcctcatctacgaggg 524
|||||
Db 167 GTCTTCCTCATCTACGAGGG 186

RESULT 7
AA644607
LOCUS
DEFINITION
5' similar to gb:XI573_cds1 6-PHOSPHOFRUCTOKINASE, LIVER TYPE
(HUMAN);, mRNA sequence.
ACCESSION
AA644607
VERSION
AA644607.1 GI:2569825
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 343)
AUTHORS
Hallier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
White,Y., Wylie,T., Waterston,R. and Wilson,R.
TITLE
WashU-NCI human EST Project
JOURNAL
Unpublished (1997)
COMMENT
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 828 Std Error: 0.00
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 336.
Location/Qualifiers
1. .343
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1047733"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"
/note="Organ: mixed (see below); Vector: pT7T3D-Pac
(Pharmacia) with a modified polylinker; Site_1: Not I;

FEATURES
source
1. .343
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1047733"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"
/note="Organ: mixed (see below); Vector: pT7T3D-Pac
(Pharmacia) with a modified polylinker; Site_1: Not I;

Site_2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NbHM, pregnant uterus
NbHPU, and fetal heart NbHH19W) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."

BASE COUNT 61 a 97 c 125 g 60 t
ORIGIN
Query Match 1.3%; Score 20; DB 9; Length 343;
Best Local Similarity 100.0%; Pred. NO. 34;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 505 gtcttcctcatctacgaggg 524
|||||
Db 122 GTCTTCCTCATCTACGAGGG 141

RESULT 8
AA629140
LOCUS
DEFINITION
AA629140 345 bp mRNA linear EST 16-OCT-1997
af50b06.s1 Soares_NhHMPu_S1 Homo sapiens cDNA clone IMAGE:1046387
3' similar to gb:XI573_cds1 6-PHOSPHOFRUCTOKINASE, LIVER TYPE
(HUMAN);, mRNA sequence.
ACCESSION
AA629140
VERSION
AA629140.1 GI:2541527
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 345)
AUTHORS
Hallier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
White,Y., Wylie,T., Waterston,R. and Wilson,R.
TITLE
WashU-NCI human EST Project
JOURNAL
Unpublished (1997)
COMMENT
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Possible reversed clone: polyT not found
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 290.
Location/Qualifiers
1. .345
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1046387"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"
/note="Organ: mixed (see below); Vector: pT7T3D-Pac
(Pharmacia) with a modified polylinker; Site_1: Not I;
Site_2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NbHM, pregnant uterus
NbHPU, and fetal heart NbHH19W) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."

FEATURES
source
1. .345
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1046387"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"
/note="Organ: mixed (see below); Vector: pT7T3D-Pac
(Pharmacia) with a modified polylinker; Site_1: Not I;
Site_2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NbHM, pregnant uterus
NbHPU, and fetal heart NbHH19W) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."

```
BASE COUNT      70 a      103 c      112 g      60 t
ORIGIN

Query Match      1.3%; Score 20; DB 9; Length 345;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 505 gtcttctctctatcagagg 524
    |||||||
Db 91 GTCTTCTCTCATCTACGAGG 110

RESULT 9
BI316879/c
LOCUS      BI316879      363 bp      mRNA      linear      EST 29-NOV-2001
DEFINITION saf77e08.y1 Gm-cl078 Glycine max cDNA clone GENOME SYSTEMS CLONE
            ID: Gm-cl078-2152 5' similar to TR:Q41668 Q41668 GUANINE NUCLEOTIDE
            REGULATORY PROTEIN. ;, mRNA sequence.
ACCESSION  BI316879
VERSION    BI316879.1 GI:14991206
KEYWORDS   EST.
SOURCE     soybean.
ORGANISM   Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE  1 (bases 1 to 363)
AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelldng,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
            Public Soybean EST Project
            Unpublished (1999)
            Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: ResGen, Invitrogen Corp. 2130
            South Memorial Parkway Huntsville, AL 35801 For further information
            call: (800)-533-4363 or contact via email: ccu@resgen.com
            Seq primer: -40RP from Gibco
            High quality sequence stop: 284.
            Location/Qualifiers
                1..363
                /organism="Glycine max"
                /db_xref="taxon:3847"
                /clone="GENOME SYSTEMS CLONE ID: Gm-cl078-2152"
                /clone_lib="Gm-cl078"
                /tissue_type="Roots of 7 day old 'Bragg' supernodulating
                mutant NTS382 seedlings"
                /dev_stage="7 days old"
                /lab_host="DH10B"
                /note="vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
                XhoI; The mRNA was isolated from roots of 7 day old
                'Bragg' supernodulating mutant NTS382 seedlings that were
                infected with Bradyrhizobium japonicum, strain USDA 110, 48
                hours prior to harvest. Dr. Gary Stacey generously donated
                the tissue. The roots were flash-frozen in liquid
                nitrogen. Stratagene's cDNA Synthesis Kit (catalog number
                200401) was used to synthesize the cDNA. First-strand
                synthesis was performed with 5-methyl dCTP, hence the
                ligated cDNA was hemimethylated. A modification of
                Stratagene's first-strand synthesis primer was used. An
                'anchor' nucleotide (V=A, C, or G) was added to the 3' end
                of the primer [GAGAGAGAGAGAGAGAGAACTACTCTCGAG(T)18V] to
                anchor the primer at the 5' end of the poly(A) tract.

FEATURES             source
                1..363
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:3062571"
                /clone_lib="NIH_MGC_38"
                /tissue_type="lymph"
                /cell_type="germinal center B cells"
                /cell_line="MGC85"
                /lab_host="DH10B (LTI)"
                /note="vector: pT7T3-Pac; Site_1: NotI; Site_2: Eco RI;
                Constructed from size fractionated cytoplasmic mRNA
                (2.5-3.5kb). Directionally cloned. Cells provided by Louis
                M. Staudt, Ph.D. Library preparation by Maria de Fatima
                Bonaldo, Ph.D. and M. Bento Soares, Ph.D."
```

```
After second-strand synthesis, the cDNA ends were filled
in with cloned Pfu DNA, ligated to EcoRI adapters and
subsequently phosphorylated. The cDNA was then
precipitated and redissolved in sterile, RNase-,
DNase-free water. The XhoI site within the first-strand
synthesis primer was then restricted by digestion with
XhoI from Promega (40U/ul); all XhoI sites in the cDNA
would be protected by their hemimethylated status. The
cDNA constructs were size-fractionated with a 500bp
cutoff, using Sephacryl S-500 High Resolution (Pharmacia
Biotech) in a 2-mm diameter column and a bed volume of
approximately 1ml. The column eluent was precipitated,
redissolved, and ligated into Stratagene's pBluescript II
XR Predigested vector (pBluescript II SK(+)) vector that
has been digested with EcoRI and XhoI, and phosphorylated
by Stratagene). This library was constructed in the
laboratory of Dr. Paul Keim and Dr. Virginia H. Coryell at
Northern Arizona University."
BASE COUNT      87 a      103 c      64 g      107 t      2 others
ORIGIN

Query Match      1.3%; Score 20; DB 10; Length 363;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 111 ggagaagaacaatttggcca 130
    |||||||
Db 241 GGAGAGAGAAACAATTGGCA 222

RESULT 10
AW407577
LOCUS      AW407577      365 bp      mRNA      linear      EST 16-FEB-2000
DEFINITION UI-HF-BM0-adp-g-02-0-UI.r1 NIH_MGC_38 Homo sapiens cDNA clone
            IMAGE:3062571 5', mRNA sequence.
ACCESSION  AW407577
VERSION    AW407577.1 GI:6926634
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 365)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Eco RI site shown at the beginning of the sequence.
            Tissue procurement: Louis M. Staudt, M.D., Ph.D.
            cDNA Library Preparation: M.B. Soares Lab
            cDNA Library Arrayed by: M.B. Soares Lab
            DNA Sequencing by: M.B. Soares Lab
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: M13 Forward.
            Location/Qualifiers
                1..365
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:3062571"
                /clone_lib="NIH_MGC_38"
                /tissue_type="lymph"
                /cell_type="germinal center B cells"
                /cell_line="MGC85"
                /lab_host="DH10B (LTI)"
                /note="vector: pT7T3-Pac; Site_1: NotI; Site_2: Eco RI;
                Constructed from size fractionated cytoplasmic mRNA
                (2.5-3.5kb). Directionally cloned. Cells provided by Louis
                M. Staudt, Ph.D. Library preparation by Maria de Fatima
                Bonaldo, Ph.D. and M. Bento Soares, Ph.D."
```

BASE COUNT 67 a 110 c 128 g 60 t
ORIGIN

Query Match 1.3%; Score 20; DB 9; Length 365;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 505 gtcttctcatctacgagg 524
|||||
Db 167 GTCTTCTCATCTACGAGG 186

RESULT 11
BM152194
LOCUS
DEFINITION
TCBAP1E6167 pediatric pre-B cell acute lymphoblastic leukemia
Baylor-HGSC project-TCBA Homo sapiens cDNA clone TCBAP6167, mRNA
sequence.

ACCESSION
BM152194
VERSION
BM152194.1 GI:17176285
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

REFERENCE
1 (bases 1 to 365)
AUTHORS
Wei, Y., Tsang, Y.T.M., Mei, G., Ku, J.M., Ali-Osman, F.R. Jr.,
Gunaratne, P.H., Muzny, D., Bouck, J., Gibbs, R.A. and Margolin, J.F.
Pediatric leukemia cDNA Sequencing Project (2001)
Unpublished (2001)
COMMENT
Contact: Dr. Judith F. Margolin
Texas Children's Cancer Center and Human Genome Sequencing Center
at Baylor College of Medicine
1102 Bates, MC3-3320 Houston, TX 77030, USA
Tel: 832-824-4536
Fax: 832-825-4038
Email: clones@txccc.org
Seq primer: M13 primer.

FEATURES
source
1. 365
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="TCBAP6167"
/clone_lib="Pediatric pre-B cell acute lymphoblastic
leukemia Baylor-HGSC project-TCBA"
/sex="male"
/tissue_type="leukopheresis"
/cell_type="pre-B cell"
/dev_stage="pediatric 2 years"
/lab_host="DH10B"
/note="Vector: lambda PSB; Site_1: BamHI; Site_2: EcoRI;
First strand cDNA was primed with an anchored
XhoI-oligo(dT) primer [5'GGAGGACTCGAGCGCGCAGGAGGAG(T)VN
3'; V=A,C,G; N=A,C,G,T] and then dG tailed. Second strand
was primed with a BamHI-dC primer
[5'AGAGAGCTCGATCCGCGCGCAATAATAATAT(C) 3'].
Double-stranded cDNA was then digested with BamHI and XhoI
and directionally cloned into the BamHI and SalI sites of
lambda PSB vector. Library was constructed by Wei Yu at RIKEN
of Japan (Carninci P, Westover A, Nishiyama Y, Ohsumi T,
Itoh M, Nagaoka S, Sasaki, Okazaki Y, Muramatsu M,
Schneider C, Hayashizaki Y, High efficiency selection of
full-length cDNA by improved biotinylated cap trapper.,
DNA Res 4: 1, 61-6, Feb 28, 1997)."

BASE COUNT 66 a 111 c 126 g 62 t
ORIGIN

Query Match 1.3%; Score 20; DB 10; Length 365;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 505 gtcttctcatctacgagg 524
|||||
Db 159 GTCTTCTCATCTACGAGG 178

RESULT 12
AW222767/c
LOCUS
DEFINITION
AW222767 tomato fruit red ripe, TAMU Lycopersicon esculentum cDNA
clone CLEN9E24, mRNA sequence.

ACCESSION
AW222767
VERSION
AW222767.1 GI:6534451
KEYWORDS
EST.
SOURCE
tomato.
ORGANISM
Lycopersicon esculentum

REFERENCE
1 (bases 1 to 391)
AUTHORS
Alcala, J., Vrebalov, J., White, R., Matern, A.L., Holt, I.E., Llang, F.,
Upton, J., Hansen, T., Craven, M.B., Bowman, C.L., Ahn, S., Ronning, C.M.,
Fraser, C.M., Martin, G.B., Tanksley, S.D. and Giovannoni, J.
Generation of ESTs from tomato fruit tissue
Unpublished (1999)
COMMENT
Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>
5 prime sequence.

FEATURES
Location/Qualifiers
1. 391
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="CLEN9E24"
/clone_lib="tomato fruit red ripe, TAMU"
/tissue_type="pericarp"
/dev_stage="red ripe (7-20 days post-breaker)"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; supplier: Giovannoni; Fruit were tagged at the
breaker stage (first sign of lycopene accumulation on the
blossom end of the fruit) and harvested 7 days
post-breaker (fully red-ripe), 10 days post breaker, and
20 days post-breaker (over-ripe). 20 day fruit which
showed external or internal signs of pathogenesis were
discarded. Fruit were cut in half and the seeds and
locules were discarded prior to freezing the pericarp."

BASE COUNT 90 a 89 c 125 g 87 t
ORIGIN

Query Match 1.3%; Score 20; DB 9; Length 391;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 78 agtggtgcgtccgagtc 97
|||||
Db 351 AGGTGTCGTCGAGTCCC 332

RESULT 13
AL120231
LOCUS
DEFINITION
AL120231
ACCESSION
AL120231
VERSION
AL120231.1 GI:5926130
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

BASE COUNT 434 bp
ORIGIN

Query Match 1.3%; Score 20; DB 9; Length 391;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 78 agtggtgcgtccgagtc 97
|||||
Db 351 AGGTGTCGTCGAGTCCC 332

RESULT 13
AL120231
LOCUS
DEFINITION
AL120231
ACCESSION
AL120231
VERSION
AL120231.1 GI:5926130
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

BASE COUNT 434 bp
ORIGIN

Query Match 1.3%; Score 20; DB 9; Length 391;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 78 agtggtgcgtccgagtc 97
|||||
Db 351 AGGTGTCGTCGAGTCCC 332

RESULT 13
AL120231
LOCUS
DEFINITION
AL120231
ACCESSION
AL120231
VERSION
AL120231.1 GI:5926130
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

BASE COUNT 434 bp
ORIGIN

Query Match 1.3%; Score 20; DB 9; Length 391;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 78 agtggtgcgtccgagtc 97
|||||
Db 351 AGGTGTCGTCGAGTCCC 332

RESULT 13
AL120231
LOCUS
DEFINITION
AL120231
ACCESSION
AL120231
VERSION
AL120231.1 GI:5926130
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

BASE COUNT 434 bp
ORIGIN

Query Match 1.3%; Score 20; DB 9; Length 391;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 78 agtggtgcgtccgagtc 97
|||||
Db 351 AGGTGTCGTCGAGTCCC 332

RESULT 13
AL120231
LOCUS
DEFINITION
AL120231
ACCESSION
AL120231
VERSION
AL120231.1 GI:5926130
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

BASE COUNT 434 bp
ORIGIN

Query Match 1.3%; Score 20; DB 9; Length 391;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 78 agtggtgcgtccgagtc 97
|||||
Db 351 AGGTGTCGTCGAGTCCC 332

RESULT 13
AL120231
LOCUS
DEFINITION
AL120231
ACCESSION
AL120231
VERSION
AL120231.1 GI:5926130
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

BASE COUNT 434 bp
ORIGIN

Query Match 1.3%; Score 20; DB 9; Length 391;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 78 agtggtgcgtccgagtc 97
|||||
Db 351 AGGTGTCGTCGAGTCCC 332

RESULT 13
AL120231
LOCUS
DEFINITION
AL120231
ACCESSION
AL120231
VERSION
AL120231.1 GI:5926130
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

BASE COUNT 434 bp
ORIGIN

Query Match 1.3%; Score 20; DB 9; Length 391;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 78 agtggtgcgtccgagtc 97
|||||
Db 351 AGGTGTCGTCGAGTCCC 332

RESULT 13
AL120231
LOCUS
DEFINITION
AL120231
ACCESSION
AL120231
VERSION
AL120231.1 GI:5926130
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

BASE COUNT 434 bp
ORIGIN

Query Match 1.3%; Score 20; DB 9; Length 391;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 78 agtggtgcgtccgagtc 97
|||||
Db 351 AGGTGTCGTCGAGTCCC 332

RESULT 13
AL120231
LOCUS
DEFINITION
AL120231
ACCESSION
AL120231
VERSION
AL120231.1 GI:5926130
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

BASE COUNT 434 bp
ORIGIN

Query Match 1.3%; Score 20; DB 9; Length 391;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 78 agtggtgcgtccgagtc 97
|||||
Db 351 AGGTGTCGTCGAGTCCC 332

RESULT 13
AL120231
LOCUS
DEFINITION
AL120231
ACCESSION
AL120231
VERSION
AL120231.1 GI:5926130
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

BASE COUNT 434 bp
ORIGIN

Query Match 1.3%; Score 20; DB 9; Length 391;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 78 agtggtgcgtccgagtc 97
|||||
Db 351 AGGTGTCGTCGAGTCCC 332

RESULT 13
AL120231
LOCUS
DEFINITION
AL120231
ACCESSION
AL120231
VERSION
AL120231.1 GI:5926130
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

BASE COUNT 434 bp
ORIGIN

Query Match 1.3%; Score 20; DB 9; Length 391;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 78 agtggtgcgtccgagtc 97
|||||
Db 351 AGGTGTCGTCGAGTCCC 332

RESULT 13
AL120231
LOCUS
DEFINITION
AL120231
ACCESSION
AL120231
VERSION
AL120231.1 GI:5926130
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

BASE COUNT 434 bp
ORIGIN

Query Match 1.3%; Score 20; DB 9; Length 391;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 78 agtggtgcgtccgagtc 97
|||||
Db 351 AGGTGTCGTCGAGTCCC 332

RESULT 13
AL120231
LOCUS
DEFINITION
AL120231
ACCESSION
AL120231
VERSION
AL120231.1 GI:5926130
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

BASE COUNT 434 bp
ORIGIN

Query Match 1.3%; Score 20; DB 9; Length 391;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 78 agtggtgcgtccgagtc 97
|||||
Db 351 AGGTGTCGTCGAGTCCC 332

RESULT 13
AL120231
LOCUS
DEFINITION
AL120231
ACCESSION
AL120231
VERSION
AL120231.1 GI:5926130
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

BASE COUNT 434 bp
ORIGIN

Query Match 1.3%; Score 20; DB 9; Length 391;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 78 agtggtgcgtccgagtc 97
|||||
Db 351 AGGTGTCGTCGAGTCCC 332

RESULT 13
AL120231
LOCUS
DEFINITION
AL120231
ACCESSION
AL120231
VERSION
AL120231.1 GI:5926130
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

BASE COUNT 434 bp
ORIGIN

Query Match 1.3%; Score 20; DB 9; Length 391;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 78 agtggtgcgtccgagtc 97
|||||
Db 351 AGGTGTCGTCGAGTCCC 332

RESULT 13
AL120231
LOCUS
DEFINITION
AL120231
ACCESSION
AL120231
VERSION
AL120231.1 GI:5926130
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

BASE COUNT 434 bp
ORIGIN

Query Match 1.3%; Score 20; DB 9; Length 391;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 78 agtggtgcgtccgagtc 97
|||||
Db 351 AGGTGTCGTCGAGTCCC 332

RESULT 13
AL120231
LOCUS
DEFINITION
AL120231
ACCESSION
AL120231
VERSION
AL120231.1 GI:5926130
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

BASE COUNT 434 bp
ORIGIN

Query Match 1.3%; Score 20; DB 9; Length 391;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 78 agtggtgcgtccgagtc 97
|||||
Db 351 AGGTGTCGTCGAGTCCC 332

RESULT 13
AL120231
LOCUS
DEFINITION
AL120231
ACCESSION
AL120231
VERSION
AL120231.1 GI:5926130
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

BASE COUNT 434 bp
ORIGIN

Query Match 1.3%; Score 20; DB 9; Length 391;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 78 agtggtgcgtccgagtc 97
|||||
Db 351 AGGTGTCGTCGAGTCCC 332

RESULT 13
AL120231
LOCUS
DEFINITION
AL120231
ACCESSION
AL120231
VERSION
AL120231.1 GI:5926130
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

BASE COUNT 434 bp
ORIGIN

Query Match 1.3%; Score 20; DB 9; Length 391;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 78 agtggtgcgtccgagtc 97
|||||
Db 351 AGGTGTCGTCGAGTCCC 332

RESULT 13
AL120231
LOCUS
DEFINITION
AL120231
ACCESSION
AL120231
VERSION
AL120231.1 GI:5926130
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

BASE COUNT 434 bp
ORIGIN

Query Match 1.3%; Score 20; DB 9; Length 391;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 78 agtggtgcgtccgagtc 97
|||||
Db 351 AGGTGTCGTCGAGTCCC 332

RESULT 13
AL120231
LOCUS
DEFINITION
AL120231
ACCESSION
AL120231
VERSION
AL120231.1 GI:5926130
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

BASE COUNT 434 bp
ORIGIN

Query Match 1.3%; Score 20; DB 9; Length 391;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 78 agtggtgcgtccgagtc 97
|||||
Db 351 AGGTGTCGTCGAGTCCC 332

RESULT 13
AL120231
LOCUS
DEFINITION
AL120231
ACCESSION
AL120231
VERSION
AL120231.1 GI:5926130
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

BASE COUNT 434 bp
ORIGIN

Query Match 1.3%; Score 20; DB 9; Length 391;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 78 agtggtgcgtccgagtc 97
|||||
Db 351 AGGTGTCGTCGAGTCCC 332

RESULT 13
AL120231
LOCUS
DEFINITION
AL120231
ACCESSION
AL120231
VERSION
AL120231.1 GI:5926130
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

BASE COUNT 434 bp
ORIGIN

Query Match 1.3%; Score 20; DB 9; Length 391;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 78 agtggtgcgtccgagtc 97
|||||
Db 351 AGGTGTCGTCGAGTCCC 332

RESULT 13
AL120231
LOCUS
DEFINITION
AL120231
ACCESSION
AL120231
VERSION
AL120231.1 GI:5926130
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

BASE COUNT 434 bp
ORIGIN

Query Match 1.3%; Score 20; DB 9; Length 391;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 78 agtggtgcgtccgagtc 97
|||||
Db 351 AGGTGTCGTCGAGTCCC 332

RESULT 13
AL120231
LOCUS
DEFINITION
AL120231
ACCESSION
AL120231
VERSION
AL120231.1 GI:5926130
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

BASE COUNT 434 bp
ORIGIN

Query Match 1.3%; Score 20; DB 9; Length 391;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 78 agtggtgcgtccgagtc 97
|||||
Db 351 AGGTGTCGTCGAGTCCC 332

RESULT 13
AL120231
LOCUS
DEFINITION
AL120231
ACCESSION
AL120231
VERSION
AL120231.1 GI:5926130
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

BASE COUNT 434 bp
ORIGIN

Query Match 1.3%; Score 20; DB 9; Length 391;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 78 agtggtgcgtccgagtc 97
|||||
Db 351 AGGTGTCGTCGAGTCCC 332

RESULT 13
AL120231
LOCUS
DEFINITION
AL120231
ACCESSION
AL120231
VERSION
AL120231.1 GI:5926130
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

BASE COUNT 434 bp
ORIGIN

Query Match 1.3%; Score 20; DB 9; Length 391;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 78 agtggtgcgtccgagtc 97
|||||
Db 351 AGGTGTCGTCGAGTCCC 332

RESULT 13
AL120231
LOCUS
DEFINITION
AL120231
ACCESSION
AL120231
VERSION
AL120231.1 GI:5926130
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

BASE COUNT 434 bp
ORIGIN

Query Match 1.3%; Score 20; DB 9; Length 391;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 78 agtggtgcgtccgagtc 97
|||||
Db 351 AGGTGTCGTCGAGTCCC 332

RESULT 13
AL120231
LOCUS
DEFINITION
AL120231
ACCESSION
AL120231
VERSION
AL120231.1 GI:5926130
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

BASE COUNT 434 bp
ORIGIN

Query Match 1.3%; Score 20; DB 9; Length 391;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 78 agtggtgcgtccgagtc 97
|||||
Db 351 AGGTGTCGTCGAGTCCC 332

RESULT 13
AL120231
LOCUS
DEFINITION
AL120231
ACCESSION
AL120231
VERSION
AL120231.1 GI:5926130
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

BASE COUNT 434 bp
ORIGIN

Query Match 1.3%; Score 20; DB 9; Length 391;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 78 agtggtgcgtccgagtc 97
|||||
Db 351 AGGTGTCGTCGAGTCCC 332

RESULT 13
AL120231
LOCUS
DEFINITION
AL120231
ACCESSION
AL120231
VERSION
AL120231.1 GI:5926130
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

BASE COUNT 434 bp
ORIGIN

Query Match 1.3%; Score 20; DB 9; Length 391;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 78 agtggtgcgtccgagtc 97
|||||
Db 351 AGGTGTCGTCGAGTCCC 332

RESULT 13
AL120231
LOCUS
DEFINITION
AL120231
ACCESSION
AL120231
VERSION
AL120231.1 GI:5926130
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

BASE COUNT 434 bp
ORIGIN

Query Match 1.3%; Score 20; DB 9; Length 391;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 78 agtggtgcgtccgagtc 97
|||||
Db 351 AGGTGTCGTCGAGTCCC 332

RESULT 13
AL120231
LOCUS
DEFINITION
AL120231
ACCESSION
AL120231
VERSION
AL120231.1 GI:5926130
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

BASE COUNT 434 bp
ORIGIN

Query Match 1.3%; Score 20; DB 9; Length 391;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 78 agtggtgcgtccgagtc 97
|||||
Db 351 AGGTGTCGTCGAGTCCC 332

RESULT 13
AL120231
LOCUS
DEFINITION
AL120231
ACCESSION
AL120231
VERSION
AL120231.1 GI:5926130
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

BASE COUNT 434 bp
ORIGIN

Query Match 1.3%; Score 20; DB 9; Length 391;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 78 agtggtgcgtccgagtc 97
|||||
Db 351 AGGTGTCGTCGAGTCCC 332

RESULT 13
AL120231
LOCUS
DEFINITION
AL120231
ACCESSION
AL120231
VERSION
AL120231.1 GI:5926130
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

BASE COUNT 434 bp
ORIGIN

Query Match 1.3%; Score 20; DB 9; Length 391;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 78 agtggtgcgtccgagtc 97
|||||
Db 351 AGGTGTCGTCGAGTCCC 332

RESULT 13
AL120231
LOCUS
DEFINITION
AL120231
ACCESSION
AL120231
VERSION
AL120231.1 GI:5926130
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

BASE COUNT 434 bp
ORIGIN

Query Match 1.3%; Score 20; DB 9; Length 391;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 78 agtggtgcgtccgagtc 97
|||||
Db 351 AGGTGTCGTCGAGTCCC 332

RESULT 13
AL120231
LOCUS
DEFINITION
AL120231
ACCESSION
AL120231
VERSION
AL120231.1 GI:5926130
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

BASE COUNT 434 bp
ORIGIN

Query Match 1.3%; Score 20; DB 9; Length 391;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 78 agtggtgcgtccgagtc 97
|||||
Db 351 AGGTGTCGTCGAGTCCC 332

RESULT 13
AL120231
LOCUS
DEFINITION
AL120231
ACCESSION
AL120231
VERSION
AL120231.1 GI:5926130
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

BASE COUNT 434 bp
ORIGIN

Query Match 1.3%; Score 20; DB 9; Length 391;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 78 agtggtgcgtccgagtc 97
|||||
Db 351 AGGTGTCGTCGAGTCCC 332

RESULT 13
AL120231
LOCUS
DEFINITION
AL120231
ACCESSION
AL120231
VERSION
AL120231.1 GI:5926130
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

BASE COUNT 434 bp
ORIGIN

Query Match 1.3%; Score 20; DB 9; Length 391;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 78 agtggtgcgtccgagtc 97
|||||
Db 351 AGGTGTCGTCGAGTCCC 332

RESULT 13
AL120231
LOCUS
DEFINITION
AL120231
ACCESSION
AL120231
VERSION
AL120231.1 GI:5926130
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

BASE COUNT 434 bp
ORIGIN

Query Match 1.3%; Score 20; DB 9; Length 391;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 78 agtggtgcgtccgagtc 97
|||||
Db 351 AGGTGTCGTCGAGTCCC 332

RESULT 13
AL120231
LOCUS
DEFINITION
AL120231
ACCESSION
AL120231
VERSION
AL120231.1 GI:5926130
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

BASE COUNT 434 bp
ORIGIN

Query Match 1.3%; Score 20; DB 9; Length 391;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 78 agtggtgcgtccgagtc 97
|||||
Db 351 AGGTGTCGTCGAGTCCC 332

RESULT 13
AL120231
LOCUS
DEFINITION
AL120231
ACCESSION
AL120231
VERSION
AL120231.1 GI:5926130
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

BASE COUNT 434 bp
ORIGIN

Query Match 1.3%; Score 20; DB 9; Length 391;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 78 agtggtgcgtccgagtc 97
|||||
Db 351 AGGTGTCGTCGAGTCCC 332

RESULT 13
AL120231
LOCUS
DEFINITION
AL120231
ACCESSION
AL120231
VERSION
AL120231.1 GI:5926130
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

BASE COUNT 434 bp
ORIGIN

Query Match 1.3%; Score 20; DB 9; Length 391;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 78 agtggtgcgtccgagtc 97
|||||
Db 351 AGGTGTCGTCGAGTCCC 332

RESULT 13
AL120231
LOCUS
DEFINITION
AL120231
ACCESSION
AL120231
VERSION
AL120231.1 GI:5926130
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

BASE COUNT 434 bp
ORIGIN

Query Match 1.3%;

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS 1 (bases 1 to 434)
TITLE Koehrer,K., Beyer,A., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
JOURNAL EST {Koehrer, et al.)
COMMENT Unpublished (1999)
Contact: Koehrer K
MIPS
Am Klopferspitz 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by BMFZ (Biomedical Research Center at the Charite, Berlin/Germany) within the cDNA sequencing consortium of the German Genome Project.
No sl sequence available.
This clone (DKFZp761G067) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES
source
1..434
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZp761G067"
/clone_lib="761 (synonym: hamy2)"
/tissue_type="amygdala"
/dev_stage="adult"
/lab_host="DH10B"
/note="vector: pSport1; Site_1: NotI; Site_2: SalI"
BASE COUNT 77 a 129 c 157 g 71 t
ORIGIN

Query Match 1.3%; Score 20; DB 9; Length 434;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 505 gtcttcctcatctacgaggg 524
|||||
Db 181 GTCTTCCTCATCTACGAGGG 200

RESULT 14
BM174982
LOCUS BM174982 455 bp mRNA linear EST 06-DEC-2001
DEFINITION TgESTzyb17c01.y1 TGRH Tachyzoite Subtracted cDNA Library Toxoplasma gondii cDNA clone TgESTzyb17c01.y1 5', mRNA sequence.
ACCESSION BM174982
VERSION BM174982.1 GI:17393200
KEYWORDS EST.
SOURCE Toxoplasma gondii.
ORGANISM Toxoplasma gondii
Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
Sarcocystidae; Toxoplasma.
REFERENCE 1 (bases 1 to 455)
AUTHORS Tang,K., Cole,R., Fogarty,S., Sibley,L.D., Ajioke,J.A., White,M., Clifton,S., Pape,D., Martin,J., Wylie,T., Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B., Bowers,Y., Gibbons,M., Ritter ,E., Bennett,J., Franklin,C., Tsagarelshvili,R., Ronko,I., Kennedy ,S., Maguire,L., Waterston,R. and Wilson,R.
Toxoplasma EST Project
Unpublished (2001)
Contact: Clifton, S.
Toxoplasma EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: toxo@watson.wustl.edu
Contact David Sibley (toxoeest@borcim.wustl.edu) for further information relating to organism, libraries, or clone availability.
Seq primer: -40Rp from Gibco
High quality sequence stop: 420.

FEATURES
source
1..455
/organism="Toxoplasma gondii"
/strain="RH (Type I)"
/db_xref="taxon:5811"
/clone="TgESTzyb17c01.y1"
/clone_lib="TGRH Tachyzoite Subtracted cDNA Library"
/dev_stage="Tachyzoite"
/lab_host="DH10B (GeneHog, Invitrogen, Inc.)"
/note="Vector: pBluscript SK-; Site_1: EcoRI; Site_2: XhoI
; Toxoplasma RH strain tachyzoites were grown in human foreskin fibroblast cultures in vitro. The library was originally constructed by K.L.Wan, Cambridge University. cDNAs were synthesized from polyA RNAs by oligo d(T) priming and directionally cloned into the EcoRI to XhoI sites of the lambda ZapII vector using the ZAP-cDNA synthesis kit (Stratagene). The primary cDNA library was mass excised as phagemid using ExAssist helper phage (Stratagene). Phagemid DNA was extracted by phenol-chloroform method, and hybridized against a pool of over-represented ESTs (N>=12, from 5596 previous reads). The subtracted library was electroporated into DH10B (GeneHog, Invitrogen, Inc). WARNING: the library contains a small percentage of cDNAs derived from the human host cells. Library Source: David Sibley, Washington University."

BASE COUNT 87 a 122 c 154 g 92 t
ORIGIN

Query Match 1.3%; Score 20; DB 10; Length 455;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1329 accgcgtggaatcggtttc 1348
|||||
Db 327 ACCGCGTGGAAATCGGTTTC 346

RESULT 15
BE241865
LOCUS BE241865 455 bp mRNA linear EST 03-OCT-2001
DEFINITION TCAAP2E0007 Pediatric acute myelogenous leukemia cell (FAB M1) Baylor-HGSC project-TCAA Homo sapiens cDNA clone TCAAP0007, mRNA sequence.
ACCESSION BE241865
VERSION BE241865.1 GI:9093579
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Cranlata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 455)
AUTHORS Wei,Y., Tsang,Y.T.M., Mei,G., Ku,J.M., Ali-Osman Jr.,F.R., Muzny,D., Bouck,J., Gibbs,R.A. and Margolin,J.F.
TITLE Pediatric Leukemia cDNA Sequencing Project
JOURNAL Unpublished (2000)
COMMENT Contact: Dr. Judith F. Margolin
Texas Children's Cancer Center and Human Genome Sequencing Center at Baylor College of Medicine
1102 Bates, MC3-3320 Houston, TX 77030, USA
Tel: 832-824-4536
Fax: 832-825-4038
Email: clones@txccc.org
Citation: Carninci,P. and Hayashizaki,Y. High efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Seq primer: M13 primer.
Location/Qualifiers
1..455
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="TCAAP0007"
/clone_lib="Pediatric acute myelogenous leukemia cell (FAB

M1) Baylor-HGSC project-TCAA"
/sex="male"
/tissue_type="leukopheresis"
/cell_type="myeloid cell"
/dev_stage="pediatric 6 years"
/lab_host="DH10B"
/note="Vector: lambda pSB; Site_1: BamHI; Site_2: EcoRI;
First strand cDNA was primed with an anchored
XhoI-oligo(dT) primer [5'GGAGGACTCGAGCGCGCAGGAGAG(T)VN
3'; V=A,C,G; N=A,C,G,T] and then dG tailed. Second strand
was primed with a BamHI-dC primer
[5'AGAGAGCTCGGATCCGGCGCGCAATAATAATAT(C) 3'].
Double-stranded cDNA was then digested with BamHI and XhoI
and directionally cloned into the BamHI and Sali sites of
lambda pSB vector. Library went through one round of
normalization. Library was constructed by Wei Yu at RIKEN
of Japan (Carninci P, Westover A, Nishiyama Y, Ohsumi T,
Itoh M, Nagaoka S, Sasaki, Okazaki Y, Muramatsu M,
Schneider C, Hayashizaki Y, High efficiency selection of
full-length cDNA by improved biotinylated cap trapper.,
DNA Res 4: 1, 61-6, Feb 28, 1997)"
BASE COUNT 83 a 133 c 164 g 75 t
ORIGIN

Query Match 1.3% Score 20; DB 10; Length 455;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 505 gtcttcctcatctacgagg 524
 |||||||||
Db 158 GTCTTCCTCATCTACGAGG 177

Search completed: April 27, 2002, 02:58:50
Job time: 13442 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2002, 07:44:44 ; Search time 95.74 Seconds
(without alignments)
3027.782 Million cell updates/sec

Title: US-09-882-694A-7
Perfect score: 1578
Sequence: 1 atggactccagacaaagtgg.....aatttcaagttgggccaatga 1578

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 139433 seqs, 91850399 residues

Word size : 0

Total number of hits satisfying chosen parameters: 278866

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending_Patents_NA_New:*
1: /cgn2_6/ptodata/1/pna/PCT_NEW_COMB.seq:*
2: /cgn2_6/ptodata/1/pna/US06_NEW_COMB.seq:*
3: /cgn2_6/ptodata/1/pna/US07_NEW_COMB.seq:*
4: /cgn2_6/ptodata/1/pna/US08_NEW_COMB.seq:*
5: /cgn2_6/ptodata/1/pna/US09_NEW_COMB.seq:*
6: /cgn2_6/ptodata/1/pna/US10_NEW_COMB.seq:*
7: /cgn2_6/ptodata/1/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17	1.1	1497	5	US-09-540-209B-796
C 2	16	1.0	209	5	US-09-540-210B-6207
C 3	16	1.0	238	5	US-09-975-254-11574
4	16	1.0	239	5	US-09-540-210B-927
C 5	16	1.0	254	5	US-09-975-254-8433
C 6	16	1.0	254	5	US-09-975-254-15910
C 7	16	1.0	260	5	US-09-540-210B-27025
9	16	1.0	271	5	US-09-975-254-25409
C 10	16	1.0	292	5	US-09-540-210B-35245
11	16	1.0	294	5	US-09-789-189-1738
C 12	16	1.0	619	6	US-10-103-313-212
C 13	16	1.0	840	5	US-09-540-209B-250
C 14	16	1.0	857	7	US-60-365-384-443
C 15	16	1.0	1004	7	US-60-365-264-87
C 16	16	1.0	1428	1	PCT-US02-07826-70
C 17	16	1.0	1428	6	US-10-097-340-70
18	16	1.0	1504	6	US-10-105-299-476
C 19	16	1.0	1972	6	US-10-105-299-7237
C 20	16	1.0	1981	6	US-10-105-299-467
C 21	16	1.0	1982	7	US-60-365-384-96
C 22	16	1.0	2220	6	US-10-105-299-2789
C 23	16	1.0	2623	6	US-10-112-699-2907
C 24	16	1.0	2710	1	PCT-US02-09921-518
C 25	16	1.0	2861	6	US-10-106-698-264
26	16	1.0	3516	6	US-10-105-299-10923

C 27	16	1.0	3673	1	PCT-US02-09921-408	Sequence 408, App
C 28	16	1.0	4281	6	US-10-116-802-153	Sequence 153, App
29	16	1.0	4713	1	PCT-US02-09944-386	Sequence 386, App
C 30	16	1.0	5185	7	US-60-365-264-401	Sequence 401, App
C 31	16	1.0	10367	6	US-10-105-299-12588	Sequence 12588, A
32	16	1.0	16347	6	US-10-105-299-6563	Sequence 6563, Ap
C 33	16	1.0	20295	6	US-10-105-299-7351	Sequence 7351, Ap
C 34	16	1.0	20295	6	US-10-105-299-7354	Sequence 7354, Ap
C 35	16	1.0	21010	6	US-10-105-299-7338	Sequence 7338, Ap
C 36	16	1.0	21010	6	US-10-105-299-7338	Sequence 7338, Ap
37	16	1.0	21024	6	US-10-105-299-7339	Sequence 7339, Ap
C 38	16	1.0	21024	6	US-10-105-299-7339	Sequence 7339, Ap
C 39	16	1.0	24413	1	PCT-US02-09239-370	Sequence 370, App
C 40	16	1.0	24413	6	US-10-105-299-7346	Sequence 7346, Ap
C 41	16	1.0	24413	6	US-10-105-299-13916	Sequence 13916, A
C 42	16	1.0	26439	6	US-10-105-299-7958	Sequence 7958, Ap
C 43	16	1.0	26439	6	US-10-105-299-12665	Sequence 12665, A
C 44	16	1.0	31328	6	US-10-109-310-19	Sequence 19, Appl
45	15	1.0	93	5	US-09-975-254-10253	Sequence 10253, A

ALIGNMENTS

RESULT 1
US-09-540-209B-796
; Sequence 796, Application US/09540209B
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES
; FILE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1001-001
; CURRENT APPLICATION NUMBER: US/09/540,209B
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 796
; LENGTH: 1497
; TYPE: DNA
; ORGANISM: B.fragilis
US-09-540-209B-796

Query Match 1.1%; Score 17; DB 5; Length 1497;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 771 cgcagcaatgcattctgg 787
|||||
Db 632 cgcagcaatgcattctgg 648

RESULT 2
US-09-540-210B-6207/c
; Sequence 6207, Application US/09540210B
; GENERAL INFORMATION:
; APPLICANT: Seilhamer, Jeffrey J.
; APPLICANT: Delegeane, Angelo M.
; APPLICANT: Stuart, Susan G.
; APPLICANT: Stuve, Laura L.
; APPLICANT: Mullahy, Sara J.
; APPLICANT: Naughton, Rebecca E.
; TITLE OF INVENTION: POLYNUCLEOTIDES OF URINARY TRACT TISSUE
; FILE REFERENCE: PD-1037 CIP
; CURRENT APPLICATION NUMBER: 2002-04-03
; CURRENT FILING DATE: 2002-04-03
; PRIOR APPLICATION NUMBER: 08/972,899
; PRIOR FILING DATE: November 18, 1997
; PRIOR APPLICATION NUMBER: 08/395,244
; PRIOR FILING DATE: February 27, 1995
; PRIOR APPLICATION NUMBER: 08/722,922
; PRIOR FILING DATE: September 27, 1996
; PRIOR APPLICATION NUMBER: 60/005,526
; PRIOR FILING DATE: September 29, 1995

```

; PRIOR APPLICATION NUMBER: 08/824,029
; PRIOR FILING DATE: March 25, 1997
; PRIOR APPLICATION NUMBER: 60/014,010
; PRIOR FILING DATE: March 25, 1996
; PRIOR APPLICATION NUMBER: 08/826,847
; PRIOR FILING DATE: April 10, 1997
; PRIOR APPLICATION NUMBER: 60/015,533
; PRIOR FILING DATE: April 10, 1996
; PRIOR APPLICATION NUMBER: 08/903,555
; PRIOR FILING DATE: July 31, 1997
; PRIOR APPLICATION NUMBER: 60/023,308
; PRIOR FILING DATE: July 31, 1996
; PRIOR APPLICATION NUMBER: 08/862,178
; PRIOR FILING DATE: May 22, 1997
; PRIOR APPLICATION NUMBER: 60/018,217
; PRIOR FILING DATE: May 23, 1996
; PRIOR APPLICATION NUMBER: 08/881,589
; PRIOR FILING DATE: June 24, 1997
; PRIOR APPLICATION NUMBER: 60/021,275
; PRIOR FILING DATE: June 25, 1996
; PRIOR APPLICATION NUMBER: 08/903,802
; PRIOR FILING DATE: July 31, 1997
; PRIOR APPLICATION NUMBER: 60/023,308
; PRIOR FILING DATE: July 31, 1996
; PRIOR APPLICATION NUMBER: 08/905,881
; PRIOR FILING DATE: August 1, 1997
; PRIOR APPLICATION NUMBER: 60/025,204
; PRIOR FILING DATE: August 1, 1996
; PRIOR APPLICATION NUMBER: 08/903,471
; PRIOR FILING DATE: July 30, 1997
; PRIOR APPLICATION NUMBER: 60/025,478
; PRIOR FILING DATE: July 31, 1996
; PRIOR APPLICATION NUMBER: 08/903,556
; PRIOR FILING DATE: July 31, 1997
; PRIOR APPLICATION NUMBER: 60/025,217
; PRIOR FILING DATE: August 22, 1996
; PRIOR APPLICATION NUMBER: 08/937,142
; PRIOR FILING DATE: September 23, 1997
; PRIOR APPLICATION NUMBER: 60/026,598
; PRIOR FILING DATE: September 24, 1996
; PRIOR APPLICATION NUMBER: 08/960,746
; PRIOR FILING DATE: October 29, 1997
; PRIOR APPLICATION NUMBER: 60/030,144
; PRIOR FILING DATE: October 30, 1996
; PRIOR APPLICATION NUMBER: 08/826,847
; PRIOR FILING DATE: April 10, 1997
; PRIOR APPLICATION NUMBER: 60/015,533
; PRIOR FILING DATE: April 10, 1996
; PRIOR APPLICATION NUMBER: 08/755,524
; PRIOR FILING DATE: November 22, 1996
; PRIOR APPLICATION NUMBER: 60/007,495
; PRIOR FILING DATE: November 22, 1995
; PRIOR APPLICATION NUMBER: 09/021,031
; PRIOR FILING DATE: February 10, 1998
; PRIOR APPLICATION NUMBER: 60/039,325
; PRIOR FILING DATE: February 13, 1997
; PRIOR APPLICATION NUMBER: 09/035,172
; PRIOR FILING DATE: March 4, 1998
; PRIOR APPLICATION NUMBER: 60/040,431
; PRIOR FILING DATE: March 5, 1997
; PRIOR APPLICATION NUMBER: 09/041,894
; PRIOR FILING DATE: March 12, 1998
; PRIOR APPLICATION NUMBER: 60/040,199
; PRIOR FILING DATE: March 14, 1997
; PRIOR APPLICATION NUMBER: 09/050,817
; PRIOR FILING DATE: March 30, 1998
; PRIOR APPLICATION NUMBER: 60/043,792
; PRIOR FILING DATE: April 11, 1997
; PRIOR APPLICATION NUMBER: 09/074,999
; PRIOR FILING DATE: May 8, 1998
; PRIOR APPLICATION NUMBER: 60/048,431
; PRIOR FILING DATE: May 29, 1997
; PRIOR APPLICATION NUMBER: 09/107,592

```

```

; PRIOR FILING DATE: June 30, 1998
; PRIOR APPLICATION NUMBER: 60/052,751
; PRIOR FILING DATE: July 1, 1997
; PRIOR APPLICATION NUMBER: 09/094,079
; PRIOR FILING DATE: June 9, 1998
; PRIOR APPLICATION NUMBER: 60/049,975
; PRIOR FILING DATE: June 13, 1997
; NUMBER OF SEQ ID NOS: 35634
; SOFTWARE: PERL Program
; SEQ ID NO 6207
; LENGTH: 209
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: hu00736504
; NAME/KEY: unsure
; LOCATION: 35, 71
; OTHER INFORMATION: a, t, c, g, or other
US-09-540-210B-6207

Query Match 1.0%; Score 16; DB 5; Length 209;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 478 cacacatactcccagg 493
   |||||
Db 203 CACACATACTCCAGG 188

RESULT 3
US-09-975-254-11574/c
; Sequence 11574, Application US/09975254
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: Heck, Gregory R.
; APPLICANT: La Rosa, Thomas J.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(15309)B
; CURRENT APPLICATION NUMBER: US/09/975,254
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US/09/263,191
; PRIOR FILING DATE: 1999-03-05
; NUMBER OF SEQ ID NOS: 31255
; SEQ ID NO 11574
; LENGTH: 238
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: 700904990H1
US-09-975-254-11574

Query Match 1.0%; Score 16; DB 5; Length 238;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 gacggcgcgccagg 81
   |||||
Db 158 GACGGCGCGGCAGGT 143

RESULT 4
US-09-540-210B-927
; Sequence 927, Application US/09540210B
; GENERAL INFORMATION:
; APPLICANT: Seilhamer, Jeffrey J.
; APPLICANT: Delegeane, Angelo M.
; APPLICANT: Stuart, Susan G.
; APPLICANT: Stuve, Laura L.
; APPLICANT: Mullahy, Sara J.
; APPLICANT: Naughton, Rebecca E.

```



```
US-09-975-254-15910/c
; Sequence 15910, Application US/09975254
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: Heck, Gregory R.
; APPLICANT: La Rosa, Thomas J.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(15309)B
; CURRENT APPLICATION NUMBER: US/09/975,254
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US/09/263,191
; PRIOR FILING DATE: 1999-03-05
; NUMBER OF SEQ ID NOS: 31255
; SEQ ID NO 15910
; LENGTH: 254
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: 700952522H1
US-09-975-254-15910

Query Match      1.0%; Score 16; DB 5; Length 254;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 gacggcgggcggcaggt 81
   |||||
DB 78 GACGGCGGGCGGCAGGT 63

RESULT 7
US-09-540-210B-27025/c
; Sequence 27025, Application US/09540210B
; GENERAL INFORMATION:
; APPLICANT: Seilhamer, Jeffrey J.
; APPLICANT: Delegeane, Angelo M.
; APPLICANT: Stuart, Susan G.
; APPLICANT: Stuve, Laura L.
; APPLICANT: Mullahy, Sara J.
; APPLICANT: Naughton, Rebecca E.
; TITLE OF INVENTION: POLYNUCLEOTIDES OF URINARY TRACT TISSUE
; FILE REFERENCE: PD-1037 CIP
; CURRENT APPLICATION NUMBER: US/09/540,210B
; CURRENT FILING DATE: 2002-04-03
; PRIOR APPLICATION NUMBER: 08/972,899
; PRIOR FILING DATE: November 18, 1997
; PRIOR APPLICATION NUMBER: 08/395,244
; PRIOR FILING DATE: February 27, 1995
; PRIOR APPLICATION NUMBER: 08/722,922
; PRIOR FILING DATE: September 27, 1996
; PRIOR APPLICATION NUMBER: 60/005,526
; PRIOR FILING DATE: September 29, 1995
; PRIOR APPLICATION NUMBER: 08/824,029
; PRIOR FILING DATE: March 25, 1997
; PRIOR APPLICATION NUMBER: 60/014,010
; PRIOR FILING DATE: March 25, 1996
; PRIOR APPLICATION NUMBER: 08/826,847
; PRIOR FILING DATE: April 10, 1997
; PRIOR APPLICATION NUMBER: 60/015,533
; PRIOR FILING DATE: April 10, 1996
; PRIOR APPLICATION NUMBER: 08/903,555
; PRIOR FILING DATE: July 31, 1997
; PRIOR APPLICATION NUMBER: 60/023,308
; PRIOR FILING DATE: July 31, 1996
; PRIOR APPLICATION NUMBER: 08/862,178
; PRIOR FILING DATE: May 22, 1997
; PRIOR APPLICATION NUMBER: 60/018,217
; PRIOR FILING DATE: May 23, 1996
; PRIOR APPLICATION NUMBER: 08/881,589
; PRIOR FILING DATE: June 24, 1997
; PRIOR APPLICATION NUMBER: 60/021,275
; PRIOR FILING DATE: June 25, 1996

; PRIOR APPLICATION NUMBER: 08/903,802
; PRIOR FILING DATE: July 31, 1997
; PRIOR APPLICATION NUMBER: 60/023,308
; PRIOR FILING DATE: July 31, 1996
; PRIOR APPLICATION NUMBER: 08/905,881
; PRIOR FILING DATE: August 1, 1997
; PRIOR APPLICATION NUMBER: 60/025,204
; PRIOR FILING DATE: August 1, 1996
; PRIOR APPLICATION NUMBER: 08/903,471
; PRIOR FILING DATE: July 30, 1997
; PRIOR APPLICATION NUMBER: 60/025,478
; PRIOR FILING DATE: July 31, 1996
; PRIOR APPLICATION NUMBER: 08/903,556
; PRIOR FILING DATE: July 31, 1997
; PRIOR APPLICATION NUMBER: 60/025,217
; PRIOR FILING DATE: August 22, 1996
; PRIOR APPLICATION NUMBER: 08/937,142
; PRIOR FILING DATE: September 23, 1997
; PRIOR APPLICATION NUMBER: 60/026,598
; PRIOR FILING DATE: September 24, 1996
; PRIOR APPLICATION NUMBER: 08/960,746
; PRIOR FILING DATE: October 29, 1997
; PRIOR APPLICATION NUMBER: 60/030,144
; PRIOR FILING DATE: October 30, 1996
; PRIOR APPLICATION NUMBER: 08/826,847
; PRIOR FILING DATE: April 10, 1997
; PRIOR APPLICATION NUMBER: 60/015,533
; PRIOR FILING DATE: April 10, 1996
; PRIOR APPLICATION NUMBER: 08/755,524
; PRIOR FILING DATE: November 22, 1996
; PRIOR APPLICATION NUMBER: 60/007,495
; PRIOR FILING DATE: November 22, 1995
; PRIOR APPLICATION NUMBER: 09/021,031
; PRIOR FILING DATE: February 10, 1998
; PRIOR APPLICATION NUMBER: 60/039,325
; PRIOR FILING DATE: February 13, 1997
; PRIOR APPLICATION NUMBER: 09/035,172
; PRIOR FILING DATE: March 4, 1998
; PRIOR APPLICATION NUMBER: 60/040,431
; PRIOR FILING DATE: March 5, 1997
; PRIOR APPLICATION NUMBER: 09/041,894
; PRIOR FILING DATE: March 12, 1998
; PRIOR APPLICATION NUMBER: 60/040,199
; PRIOR FILING DATE: March 14, 1997
; PRIOR APPLICATION NUMBER: 09/050,817
; PRIOR FILING DATE: March 30, 1998
; PRIOR APPLICATION NUMBER: 60/043,792
; PRIOR FILING DATE: April 11, 1997
; PRIOR APPLICATION NUMBER: 09/074,999
; PRIOR FILING DATE: May 8, 1998
; PRIOR APPLICATION NUMBER: 60/048,431
; PRIOR FILING DATE: May 29, 1997
; PRIOR APPLICATION NUMBER: 09/107,592
; PRIOR FILING DATE: June 30, 1998
; PRIOR APPLICATION NUMBER: 60/052,751
; PRIOR FILING DATE: July 1, 1997
; PRIOR APPLICATION NUMBER: 09/094,079
; PRIOR FILING DATE: June 9, 1998
; PRIOR APPLICATION NUMBER: 60/049,975
; PRIOR FILING DATE: June 13, 1997
; NUMBER OF SEQ ID NOS: 35654
; SOFTWARE: PERL Program
; SEQ ID NO 27025
; LENGTH: 260
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: hu00786148
; NAME/KEY: unsure
; LOCATION: 31, 136
; OTHER INFORMATION: a, t, c, g, or other
US-09-540-210B-27025
```

Query Match 1.0%; Score 16; DB 5; Length 260;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1133 ctgtttggtctctatt 1148
Db 57 CTGTTGGTCTCTATT 42

RESULT 8

US-09-975-254-25409
; Sequence 25409, Application US/09975254
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: Heck, Gregory R.
; APPLICANT: La Rosa, Thomas J.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(15309)B
; CURRENT APPLICATION NUMBER: US/09/975,254
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US/09/263,191
; PRIOR FILING DATE: 1999-03-05
; NUMBER OF SEQ ID NOS: 31255
; SEQ ID NO 25409
; LENGTH: 271
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: 700964702H1
US-09-975-254-25409

Query Match 1.0%; Score 16; DB 5; Length 271;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1246 ctctccttctctgatcc 1261
Db 154 ctctccttctctgatcc 169

RESULT 9

US-09-540-210B-26631
; Sequence 26631, Application US/09540210B
; GENERAL INFORMATION:
; APPLICANT: Seilhamer, Jeffrey J.
; APPLICANT: Delegeane, Angelo M.
; APPLICANT: Stuart, Susan G.
; APPLICANT: Stuve, Laura L.
; APPLICANT: Mullahy, Sara J.
; APPLICANT: Naughton, Rebecca E.
; TITLE OF INVENTION: POLYNUCLEOTIDES OF URINARY TRACT TISSUE
; FILE REFERENCE: PD-1037 CIP
; CURRENT APPLICATION NUMBER: US/09/540,210B
; CURRENT FILING DATE: 2002-04-03
; PRIOR APPLICATION NUMBER: 08/972,899
; PRIOR FILING DATE: November 18, 1997
; PRIOR APPLICATION NUMBER: 08/395,244
; PRIOR FILING DATE: February 27, 1995
; PRIOR APPLICATION NUMBER: 08/722,922
; PRIOR FILING DATE: September 27, 1996
; PRIOR APPLICATION NUMBER: 60/005,526
; PRIOR FILING DATE: September 29, 1995
; PRIOR APPLICATION NUMBER: 08/824,029
; PRIOR FILING DATE: March 25, 1997
; PRIOR APPLICATION NUMBER: 60/014,010
; PRIOR FILING DATE: March 25, 1996
; PRIOR APPLICATION NUMBER: 08/826,847
; PRIOR FILING DATE: April 10, 1997
; PRIOR APPLICATION NUMBER: 60/015,533
; PRIOR FILING DATE: April 10, 1996

; PRIOR APPLICATION NUMBER: 08/903,555
; PRIOR FILING DATE: July 31, 1997
; PRIOR APPLICATION NUMBER: 60/023,308
; PRIOR FILING DATE: July 31, 1996
; PRIOR APPLICATION NUMBER: 08/862,178
; PRIOR FILING DATE: May 22, 1997
; PRIOR APPLICATION NUMBER: 60/018,217
; PRIOR FILING DATE: May 23, 1996
; PRIOR APPLICATION NUMBER: 08/881,589
; PRIOR FILING DATE: June 24, 1997
; PRIOR APPLICATION NUMBER: 60/021,275
; PRIOR FILING DATE: June 25, 1996
; PRIOR APPLICATION NUMBER: 08/903,802
; PRIOR FILING DATE: July 31, 1997
; PRIOR APPLICATION NUMBER: 60/023,308
; PRIOR FILING DATE: July 31, 1996
; PRIOR APPLICATION NUMBER: 08/905,881
; PRIOR FILING DATE: August 1, 1997
; PRIOR APPLICATION NUMBER: 60/025,204
; PRIOR FILING DATE: August 1, 1996
; PRIOR APPLICATION NUMBER: 08/903,471
; PRIOR FILING DATE: July 30, 1997
; PRIOR APPLICATION NUMBER: 60/025,478
; PRIOR FILING DATE: July 31, 1996
; PRIOR APPLICATION NUMBER: 08/903,556
; PRIOR FILING DATE: July 31, 1997
; PRIOR APPLICATION NUMBER: 60/025,217
; PRIOR FILING DATE: August 22, 1996
; PRIOR APPLICATION NUMBER: 08/937,142
; PRIOR FILING DATE: September 23, 1997
; PRIOR APPLICATION NUMBER: 60/026,598
; PRIOR FILING DATE: September 24, 1996
; PRIOR APPLICATION NUMBER: 08/960,746
; PRIOR FILING DATE: October 29, 1997
; PRIOR APPLICATION NUMBER: 60/030,144
; PRIOR FILING DATE: October 30, 1996
; PRIOR APPLICATION NUMBER: 08/826,847
; PRIOR FILING DATE: April 10, 1997
; PRIOR APPLICATION NUMBER: 60/015,533
; PRIOR FILING DATE: April 10, 1996
; PRIOR APPLICATION NUMBER: 08/755,524
; PRIOR FILING DATE: November 22, 1996
; PRIOR APPLICATION NUMBER: 60/007,495
; PRIOR FILING DATE: November 22, 1995
; PRIOR APPLICATION NUMBER: 09/021,031
; PRIOR FILING DATE: February 10, 1998
; PRIOR APPLICATION NUMBER: 60/039,325
; PRIOR FILING DATE: February 13, 1997
; PRIOR APPLICATION NUMBER: 09/035,172
; PRIOR FILING DATE: March 4, 1998
; PRIOR APPLICATION NUMBER: 60/040,431
; PRIOR FILING DATE: March 5, 1997
; PRIOR APPLICATION NUMBER: 09/041,894
; PRIOR FILING DATE: March 12, 1998
; PRIOR APPLICATION NUMBER: 60/040,199
; PRIOR FILING DATE: March 14, 1997
; PRIOR APPLICATION NUMBER: 09/050,817
; PRIOR FILING DATE: March 30, 1998
; PRIOR APPLICATION NUMBER: 60/043,792
; PRIOR FILING DATE: April 11, 1997
; PRIOR APPLICATION NUMBER: 09/074,999
; PRIOR FILING DATE: May 8, 1998
; PRIOR APPLICATION NUMBER: 60/048,431
; PRIOR FILING DATE: May 29, 1997
; PRIOR APPLICATION NUMBER: 09/107,592
; PRIOR FILING DATE: June 30, 1998
; PRIOR APPLICATION NUMBER: 60/052,751
; PRIOR FILING DATE: July 1, 1997
; PRIOR APPLICATION NUMBER: 09/094,079
; PRIOR FILING DATE: June 9, 1998
; PRIOR APPLICATION NUMBER: 60/049,975
; PRIOR FILING DATE: June 13, 1997
; NUMBER OF SEQ ID NOS: 35654

; GENERAL INFORMATION:
; APPLICANT: Lelias, Jean-Michel
; TITLE OF INVENTION: Human Polynucleotides and Polypeptides
; FILE REFERENCE: 25436/1720
; CURRENT APPLICATION NUMBER: US/09/789,189
; PRIOR FILING DATE: 2001-02-20
; CURRENT APPLICATION NUMBER: 60/183452
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 2005
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1738
; LENGTH: 294
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-789-189-1738

Query Match 1.0%; Score 16; DB 5; Length 294;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 348 accaaatcaatcaat 363
|||||
Db 1 accaaatcaatcaat 16

RESULT 12

US-10-103-313-212
; Sequence 212, Application US/10103313

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; FILE REFERENCE: PJZ07C1

; CURRENT APPLICATION NUMBER: US/10/103,313

; CURRENT FILING DATE: 2002-03-12

; NUMBER OF SEQ ID NOS: 653

; Prior Application removed - See File Wrapper or Palm

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 212

; LENGTH: 619

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: (610)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: misc_feature

; LOCATION: (618)

; OTHER INFORMATION: n equals a,t,g, or c

US-10-103-313-212

Query Match 1.0%; Score 16; DB 6; Length 619;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1397 tcagctccactgac 1412
|||||
Db 352 tcagctccactgac 367

RESULT 13

US-09-540-209B-250/c

; Sequence 250, Application US/09540209B

; GENERAL INFORMATION:

; APPLICANT: Gary L. Breton

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTERIOIDES FRA

; FILE REFERENCE: 2709.1001-001

; CURRENT APPLICATION NUMBER: US/09/540,209B

; CURRENT FILING DATE: 2000-04-04

; NUMBER OF SEQ ID NOS: 10444

; SEQ ID NO 250

; LENGTH: 840
; TYPE: DNA
; ORGANISM: B.fragilis
US-09-540-209B-250

Query Match 1.0%; Score 16; DB 5; Length 840;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1130 ttctgtttgtctct 1145
|||||
Db 522 TTCTGTGTGTCCT 507

RESULT 14

US-60-365-384-443/c

; Sequence 443, Application US/60365384

; GENERAL INFORMATION:

; APPLICANT: Tang, Y. Tom

; APPLICANT: Wang, Jian-rui

; APPLICANT: Wang, Zhiwei

; APPLICANT: Zhang, Jie

; APPLICANT: Zhou, Ping

; APPLICANT: Wehrman, Tom

; APPLICANT: Wang, Jian-Rui

; APPLICANT: Ghosh, Malabika

; APPLICANT: Zhao, Qing A.

; APPLICANT: Asundi, Vinod

; APPLICANT: Ren, Feiyan

; APPLICANT: Xue, Aidong J.

; APPLICANT: Ma, Yunqing

; APPLICANT: Wang, Dunrui

; APPLICANT: Weng, Gezhi

; TITLE OF INVENTION: Novel Nucleic Acids and

; FILE REFERENCE: Polypeptides

; FILE REFERENCE: 814

; CURRENT APPLICATION NUMBER: US/60/365,384

; CURRENT FILING DATE: 2002-03-14

; NUMBER OF SEQ ID NOS: 666

; SOFTWARE: pt_FL_genes Version 6.0

; SEQ ID NO 443

; LENGTH: 857

; TYPE: DNA

; ORGANISM: Homo sapiens

US-60-365-384-443

Query Match 1.0%; Score 16; DB 7; Length 857;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 488 cccaggattcgtgca 503
|||||
Db 465 CCCAGGATCGTGGCA 450

RESULT 15

US-60-365-264-87/c

; Sequence 87, Application US/60365264

; GENERAL INFORMATION:

; APPLICANT: Tang, Y. Tom

; APPLICANT: Wang, Jian-rui

; APPLICANT: Wang, Zhiwei

; APPLICANT: Wang, Jian-Rui

; APPLICANT: Zhang, Jie

; APPLICANT: Ren, Feiyan

; APPLICANT: Zhou, Ping

; APPLICANT: Ma, Yunqing

; APPLICANT: Ghosh, Malabika

; APPLICANT: Xue, Aidong J.

; APPLICANT: Asundi, Vinod

; APPLICANT: Zhao, Qing A.

; APPLICANT: Wang, Dunrui
; APPLICANT: Goodrich, Ryle W.
; APPLICANT: Chen, Rui-hong
; APPLICANT: Wehrman, Tom
; APPLICANT: Weng, Gezh1
; TITLE OF INVENTION: Novel Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 816
; CURRENT APPLICATION NUMBER: US/60/365,264
; CURRENT FILING DATE: 2002-03-14
; NUMBER OF SEQ ID NOS: 728
; SOFTWARE: pt_FL_genes Version 6.0
; SEQ ID NO 87
; LENGTH: 1004
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (243)..(1004)
US-60-365-264-87

Query Match 1.0%; Score 16; DB 7; Length 1004;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 502 catgtcttcctcatct 517
|||||
Db 530 CATGCTTCCTCATCT 515

Search completed: April 27, 2002, 07:44:48
Job time: 17140 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2002, 07:42:26 ; Search time 8172.53 Seconds
(without alignments)
4177.393 Million cell updates/sec

Title: US-09-882-694A-7
Perfect score: 1578
Sequence: 1 atggactccagaccagtggtg.....aatttcaagttgggcatga 1578

Scoring table: OLGUO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 21979536 seqs, 1081749327 residues

Word size : 0

Total number of hits satisfying chosen parameters: 43959072

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending_Patents_NA_Main:*

1: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq:*

2: /cgn2_6/ptodata/2/pna/US06_COMB.seq:*

3: /cgn2_6/ptodata/2/pna/US07_COMB.seq:*

4: /cgn2_6/ptodata/2/pna/US080_COMB.seq:*

5: /cgn2_6/ptodata/2/pna/US081_COMB.seq:*

6: /cgn2_6/ptodata/2/pna/US082_COMB.seq:*

7: /cgn2_6/ptodata/2/pna/US083_COMB.seq:*

8: /cgn2_6/ptodata/2/pna/US084_COMB.seq:*

9: /cgn2_6/ptodata/2/pna/US085_COMB.seq:*

10: /cgn2_6/ptodata/2/pna/US086_COMB.seq:*

11: /cgn2_6/ptodata/2/pna/US087_COMB.seq:*

12: /cgn2_6/ptodata/2/pna/US088_COMB.seq:*

13: /cgn2_6/ptodata/2/pna/US089_COMB.seq:*

14: /cgn2_6/ptodata/2/pna/US090_COMB.seq:*

15: /cgn2_6/ptodata/2/pna/US091_COMB.seq:*

16: /cgn2_6/ptodata/2/pna/US092_COMB.seq:*

17: /cgn2_6/ptodata/2/pna/US093_COMB.seq:*

18: /cgn2_6/ptodata/2/pna/US094_COMB.seq:*

19: /cgn2_6/ptodata/2/pna/US095A_COMB.seq:*

20: /cgn2_6/ptodata/2/pna/US095B_COMB.seq:*

21: /cgn2_6/ptodata/2/pna/US095C_COMB.seq:*

22: /cgn2_6/ptodata/2/pna/US095D_COMB.seq:*

23: /cgn2_6/ptodata/2/pna/US096A_COMB.seq:*

24: /cgn2_6/ptodata/2/pna/US096B_COMB.seq:*

25: /cgn2_6/ptodata/2/pna/US096C_COMB.seq:*

26: /cgn2_6/ptodata/2/pna/US096D_COMB.seq:*

27: /cgn2_6/ptodata/2/pna/US096E_COMB.seq:*

28: /cgn2_6/ptodata/2/pna/US097A_COMB.seq:*

29: /cgn2_6/ptodata/2/pna/US097C_COMB.seq:*

30: /cgn2_6/ptodata/2/pna/US097C_COMB.seq:*

31: /cgn2_6/ptodata/2/pna/US098A_COMB.seq:*

32: /cgn2_6/ptodata/2/pna/US098B_COMB.seq:*

33: /cgn2_6/ptodata/2/pna/US098C_COMB.seq:*

34: /cgn2_6/ptodata/2/pna/US099A_COMB.seq:*

35: /cgn2_6/ptodata/2/pna/US099B_COMB.seq:*

36: /cgn2_6/ptodata/2/pna/US099C_COMB.seq:*

37: /cgn2_6/ptodata/2/pna/US100_COMB.seq:*

38: /cgn2_6/ptodata/2/pna/US101_COMB.seq:*

39: /cgn2_6/ptodata/2/pna/US6000_COMB.seq:*

40: /cgn2_6/ptodata/2/pna/US6001_COMB.seq:*

41: /cgn2_6/ptodata/2/pna/US6002_COMB.seq:*

42: /cgn2_6/ptodata/2/pna/US6003_COMB.seq:*

43: /cgn2_6/ptodata/2/pna/US6004_COMB.seq:*

44: /cgn2_6/ptodata/2/pna/US6005_COMB.seq:*

45: /cgn2_6/ptodata/2/pna/US6006_COMB.seq:*

46: /cgn2_6/ptodata/2/pna/US6007_COMB.seq:*

47: /cgn2_6/ptodata/2/pna/US6008_COMB.seq:*

48: /cgn2_6/ptodata/2/pna/US6009_COMB.seq:*

49: /cgn2_6/ptodata/2/pna/US6010_COMB.seq:*

50: /cgn2_6/ptodata/2/pna/US6011_COMB.seq:*

51: /cgn2_6/ptodata/2/pna/US6012_COMB.seq:*

52: /cgn2_6/ptodata/2/pna/US6013_COMB.seq:*

53: /cgn2_6/ptodata/2/pna/US6014_COMB.seq:*

54: /cgn2_6/ptodata/2/pna/US6015_COMB.seq:*

55: /cgn2_6/ptodata/2/pna/US6016_COMB.seq:*

56: /cgn2_6/ptodata/2/pna/US6017_COMB.seq:*

57: /cgn2_6/ptodata/2/pna/US6018_COMB.seq:*

58: /cgn2_6/ptodata/2/pna/US6019_COMB.seq:*

59: /cgn2_6/ptodata/2/pna/US6020_COMB.seq:*

60: /cgn2_6/ptodata/2/pna/US6021_COMB.seq:*

61: /cgn2_6/ptodata/2/pna/US6022_COMB.seq:*

62: /cgn2_6/ptodata/2/pna/US6023_COMB.seq:*

63: /cgn2_6/ptodata/2/pna/US6024_COMB.seq:*

64: /cgn2_6/ptodata/2/pna/US6025_COMB.seq:*

65: /cgn2_6/ptodata/2/pna/US6026_COMB.seq:*

66: /cgn2_6/ptodata/2/pna/US6027_COMB.seq:*

67: /cgn2_6/ptodata/2/pna/US6028_COMB.seq:*

68: /cgn2_6/ptodata/2/pna/US6029_COMB.seq:*

69: /cgn2_6/ptodata/2/pna/US6030_COMB.seq:*

70: /cgn2_6/ptodata/2/pna/US6031_COMB.seq:*

71: /cgn2_6/ptodata/2/pna/US6032_COMB.seq:*

72: /cgn2_6/ptodata/2/pna/US6033_COMB.seq:*

73: /cgn2_6/ptodata/2/pna/US6034_COMB.seq:*

74: /cgn2_6/ptodata/2/pna/US6035_COMB.seq:*

75: /cgn2_6/ptodata/2/pna/US6036_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1578	100.0	1578	17	US-09-351-224E-7	Sequence 7, Appli
2	1578	100.0	1578	26	US-09-677-488A-7	Sequence 7, Appli
3	1578	100.0	1578	26	US-09-677-682A-7	Sequence 7, Appli
4	1578	100.0	1578	26	US-09-677-682B-7	Sequence 7, Appli
5	1578	100.0	1578	33	US-09-882-694-7	Sequence 7, Appli
6	1578	100.0	1578	33	US-09-882-694A-7	Sequence 7, Appli
7	1578	100.0	1764	17	US-09-351-224E-6	Sequence 6, Appli
8	1578	100.0	1764	26	US-09-677-488A-6	Sequence 6, Appli
9	1578	100.0	1764	26	US-09-677-682A-6	Sequence 6, Appli
10	1578	100.0	1764	26	US-09-677-682B-6	Sequence 6, Appli
11	1578	100.0	1764	33	US-09-882-694-6	Sequence 6, Appli
12	1578	100.0	1764	33	US-09-882-694A-6	Sequence 6, Appli
13	937	59.4	1578	17	US-09-351-224-7	Sequence 7, Appli
14	937	59.4	1578	17	US-09-351-823-7	Sequence 7, Appli
15	937	59.4	1578	26	US-09-677-488-7	Sequence 7, Appli
16	937	59.4	1578	26	US-09-677-682-7	Sequence 7, Appli
17	937	59.4	1764	17	US-09-351-224-6	Sequence 6, Appli
18	937	59.4	1764	17	US-09-351-823-6	Sequence 6, Appli
19	937	59.4	1764	26	US-09-677-488-6	Sequence 6, Appli
20	937	59.4	1764	26	US-09-677-682-6	Sequence 6, Appli
21	21	1.3	29713	62	US-60-230-435-897	Sequence 897, App
22	21	1.3	32768	59	US-60-207-315-186	Sequence 186, App
23	21	1.3	38059	1	PCT-US01-12012-1	Sequence 1, Appli
24	21	1.3	38059	1	PCT-US01-12012-49	Sequence 49, Appli
25	21	1.3	38059	17	US-09-328-925-4	Sequence 4, Appli
26	21	1.3	38059	33	US-09-880-107-2125	Sequence 2125, Ap
27	21	1.3	38059	37	US-10-032-084-4	Sequence 4, Appli
28	21	1.3	38059	73	US-60-343-185-102	Sequence 102, App
29	20	1.3	166	28	US-09-704-424-5306	Sequence 5306, Ap
30	20	1.3	221	14	US-09-016-866-1007	Sequence 1007, Ap
31	20	1.3	221	20	US-09-532-315-12249	Sequence 12249, A

```
32      20      1.3      221      42      US-60-035-830-1007      Sequence 1007, Ap
33      20      1.3      230      20      US-09-532-315-12256      Sequence 12256, A
34      20      1.3      237      20      US-09-532-315-7117      Sequence 7117, Ap
35      20      1.3      239      13      US-08-936-937-2270      Sequence 2270, Ap
36      20      1.3      239      20      US-09-532-315-12244      Sequence 12244, A
37      20      1.3      242      14      US-09-016-866-116      Sequence 116, App
38      20      1.3      242      20      US-09-532-315-12315      Sequence 12315, A
39      20      1.3      242      42      US-60-035-830-116      Sequence 116, App
40      20      1.3      248      15      US-09-105-427-172      Sequence 172, App
41      20      1.3      248      20      US-09-532-315-12240      Sequence 12240, A
42      20      1.3      258      13      US-08-992-625-613      Sequence 613, App
43      20      1.3      258      20      US-09-532-315-12261      Sequence 12261, A
44      20      1.3      275      14      US-09-073-079-1295      Sequence 1295, Ap
45      20      1.3      275      20      US-09-532-315-12297      Sequence 12297, A
```

ALIGNMENTS

```
RESULT 1
US-09-351-224E-7
; Sequence 7, Application US/09351224E
; GENERAL INFORMATION:
; APPLICANT: Duwick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Grasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; TITLE OF INVENTION: Detoxification
; FILE REFERENCE: 5718-111
; CURRENT APPLICATION NUMBER: US/09/351.224E
; CURRENT FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows version 4.0
; SEQ ID NO 7
; LENGTH: 1578
; TYPE: DNA
; ORGANISM: Exophiala spinifera
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: {0}...(0)
; OTHER INFORMATION: permease, fully spliced cDNA
US-09-351-224E-7
```

```
Query Match      100.0%; Score 1578; DB 17; Length 1578;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1578; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 atggactccagaccagtgatagcgagagaaagcgaggacagcagacagaaagac 60
   |||||
Db 1 atggactccagaccagtgatagcgagagaaagcgaggacagcagacagaaagac 60

Qy 61 acagagacggcgcgaggtggtgcgtccagagtcacctgaacgttccctggagagaaa 120
   |||||
Db 61 acagagacggcgcgaggtggtgcgtccagagtcacctgaacgttccctggagagaaa 120

Qy 121 caattggcaccatcacatcggtgcgtccagagtcacctgaacgttccctggagagaaa 180
   |||||
Db 121 caattggcaccatcacatcggtgcgtccagagtcacctgaacgttccctggagagaaa 180

Qy 181 atccaggcagtcctcagctcgccctactagcgggggggcccgctcactctccttacggc 240
   |||||
Db 181 atccaggcagtcctcagctcgccctactagcgggggggcccgctcactctccttacggc 240

Qy 241 atcctaatacagtaactcgtctacatctgcatcgcttcttcattagccgaactgaccagc 300
   |||||
Db 241 atcctaatacagtaactcgtctacatctgcatcgcttcttcattagccgaactgaccagc 300

Qy 301 gtctacccgactgcgggtggcccaataatctttgcgtcgatccttgccacccaaataatc 360
   |||||
Db 301 gtctacccgactgcgggtggcccaataatctttgcgtcgatccttgccacccaaataatc 360
```

```
Qy 361 aatcggagcatcttcatacgtgtgcgagactcgtgctgttgccttcacgacgctatcgga 420
   |||||
Db 361 aatcggagcatcttcatacgtgtgcgagactcgtgctgttgccttcacgacgctatcgga 420

Qy 421 agtcagtgaccatgatacctgctcaacagatcccgcgctgatacgcgcctatagtcaac 480
   |||||
Db 421 agtcagtgaccatgatacctgctcaacagatcccgcgctgatacgcgcctatagtcaac 480

Qy 481 acatactcccaggtatcgtggcatgttctcctcatctacagggagtcgcgctggtgtg 540
   |||||
Db 481 acatactcccaggtatcgtggcatgttctcctcatctacagggagtcgcgctggtgtg 540

Qy 541 ctcttgttcaactgtttgcccctgaaagaaacccctgggttcatgaaatcggtatcggc 600
   |||||
Db 541 ctcttgttcaactgtttgcccctgaaagaaacccctgggttcatgaaatcggtatcggc 600

Qy 601 ctcaagatcgctctcttcgtgatctcctttatcgccattctagcgcggtccaaaccccaag 660
   |||||
Db 601 ctcaagatcgctctcttcgtgatctcctttatcgccattctagcgcggtccaaaccccaag 660

Qy 661 gctccaaaactcaacaggtatggactgcttgaggacaactatactggtggtccgacggtc 720
   |||||
Db 661 gctccaaaactcaacaggtatggactgcttgaggacaactatactggtggtccgacggtc 720

Qy 721 tgcttcactcctggcccttctcgacatcctgcttcctcatgttcattggttggacgagcaatg 780
   |||||
Db 721 tgcttcactcctggcccttctcgacatcctgcttcctcatgttcattggttggacgagcaatg 780

Qy 781 catctggctgaagaatgcacagatgctgctcgtacggtaccccaagcagtggtcagtgca 840
   |||||
Db 781 catctggctgaagaatgcacagatgctgctcgtacggtaccccaagcagtggtcagtgca 840

Qy 841 atcataaattggcttctgcaccgcttcccatatacacaatcgagttctgtatggaattaca 900
   |||||
Db 841 atcataaattggcttctgcaccgcttcccatatacacaatcgagttctgtatggaattaca 900

Qy 901 gatctcgactctattctaagtctccgctgcttatattccattcgcagacaatgacgcagtc 960
   |||||
Db 901 gatctcgactctattctaagtctccgctgcttatattccattcgcagacaatgacgcagtc 960

Qy 961 ctctggctcgtcagtttgcacaggtcctctcctcgtatggtggtggtggtggtggtggt 1020
   |||||
Db 961 ctctggctcgtcagtttgcacaggtcctctcctcgtatggtggtggtggtggtggtggt 1020

Qy 1021 gccctcaacgctgtacaagagactgctctcgactcactgagcttctgcccgggacaaat 1080
   |||||
Db 1021 gccctcaacgctgtacaagagactgctctcgactcactgagcttctgcccgggacaaat 1080

Qy 1081 gggctggtattttccactcactcgaacgcattcctcccgctggcaagtctctgttgg 1140
   |||||
Db 1081 gggctggtattttccactcactcgaacgcattcctcccgctggcaagtctctgttgg 1140

Qy 1141 tctctattcgcgacctggggaaattctggccacatgctgggatgtatatcttcaggttctagc 1200
   |||||
Db 1141 tctctattcgcgacctggggaaattctggccacatgctgggatgtatatcttcaggttctagc 1200

Qy 1201 acagctttcaaatgcttggtaaatctccgcggttgtaactccagcaactcctcctcctgac 1260
   |||||
Db 1201 acagctttcaaatgcttggtaaatctccgcggttgtaactccagcaactcctcctcctgac 1260

Qy 1261 ccaatcgccctactcctctacaaaagcgagatccaaagtcttctgcccagcactcgtgct 1320
   |||||
Db 1261 ccaatcgccctactcctctacaaaagcgagatccaaagtcttctgcccagcactcgtgct 1320

Qy 1321 tttgtgtaccgctggtggaatcgggtttctgtgtaagtgtgtagcgtggtgtcttcacgtcc 1380
   |||||
Db 1321 tttgtgtaccgctggtggaatcgggtttctgtgtaagtgtgtagcgtggtgtcttcacgtcc 1380

Qy 1381 gtcacacactgtgtttttcagcttcccactgaccgtgcctacggcgctcaacccatgaat 1440
   |||||
Db 1381 gtcacacactgtgtttttcagcttcccactgaccgtgcctacggcgctcaacccatgaat 1440
```


QY 1441 tacacaagtgcgattatagcggttgacattgctcttggtgtcttgaactgggtcgtgcat 1500
Db 1441 tacacaagtgcgattatagcggttgacattgctcttggtgtcttgaactgggtcgtgcat 1500
QY 1501 gccaggaagcattatcaggagaccccccacttggagcttgacggacgggtcgtcggagcagaa 1560
Db 1501 gccaggaagcattatcaggagaccccccacttggagcttgacggacgggtcgtcggagcagaa 1560
QY 1561 ttccaagtgggcccata 1578
Db 1561 ttccaagtgggcccata 1578

RESULT 2
US-09-677-488A-7
; Sequence 7, Application US/09677488A
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; TITLE OF INVENTION: Detoxification
; FILE REFERENCE: 35718/204100
; CURRENT APPLICATION NUMBER: US/09/677,488A
; CURRENT FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1578
; TYPE: DNA
; ORGANISM: Exophiala spinifera
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: permealase, fully spliced cDNA
US-09-677-488A-7

Query Match 100.0%; Score 1578; DB 26; Length 1578;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1578; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggactccagaccagtggtgatacggcgagaaaggcggaagggcagacaaacgaagaac 60
Db 1 atggactccagaccagtggtgatacggcgagaaaggcggaagggcagacaaacgaagaac 60
QY 61 acagagacggcgggcgaggtggtgcgtccgagtccttgaacgttccctcgtggagagaaa 120
Db 61 acagagacggcgggcgaggtggtgcgtccgagtccttgaacgttccctcgtggagagaaa 120
QY 121 caattggcaccatcacccatcgtgtcccttggccttggatgttgcaacagttgggtggt 180
Db 121 caattggcaccatcacccatcgtgtcccttggccttggatgttgcaacagttgggtggt 180
QY 181 atctcaggcagtcctcagctcgccctactagcgggggggcccgctcactctcctttacggc 240
Db 181 atctcaggcagtcctcagctcgccctactagcgggggggcccgctcactctcctttacggc 240
QY 241 atcctaatacagtaactcgtctacatctgcacgtcttctcattagccgaactgaccagc 300
Db 241 atcctaatacagtaactcgtctacatctgcacgtcttctcattagccgaactgaccagc 300
QY 301 gtctacccgactgcgggtggcccaatatcttttgcgtcgatcctggcaccacaaatcaatc 360
Db 301 gtctacccgactgcgggtggcccaatatcttttgcgtcgatcctggcaccacaaatcaatc 360
QY 361 aatcggagcatttcatacagtggtgcgggactcgtgtcgttgcgttgcgttatcggtatcgga 420
Db 361 aatcggagcatttcatacagtggtgcgggactcgtgtcgttgcgttgcgttatcggtatcgga 420

QY 421 agtcagtgaccatgatacctgtctcaacagatcccgcgctgatagccgcctatagtcac 480
Db 421 agtcagtgaccatgatacctgtctcaacagatcccgcgctgatagccgcctatagtcac 480
QY 481 acatactcccagattcgtggcatgtcttccctcatctacgagggagtcgctgggtggtg 540
Db 481 acatactcccagattcgtggcatgtcttccctcatctacgagggagtcgctgggtggtg 540
QY 541 ctcttgttcaactgtttgccctgaaaagaaacccctgggttcattgaaatcggttcggc 600
Db 541 ctcttgttcaactgtttgccctgaaaagaaacccctgggttcattgaaatcggttcggc 600
QY 601 ctacgagtcgtctcttctgtgatctcttcttattgcccattctagcgcgggtccaaaccccaag 660
Db 601 ctacgagtcgtctcttctgtgatctcttcttattgcccattctagcgcgggtccaaaccccaag 660
QY 661 gctccaaactcacaggtatggactgcttggagcaactatactggtggtccgacggcgtc 720
Db 661 gctccaaactcacaggtatggactgcttggagcaactatactggtggtccgacggcgtc 720
QY 721 tgcttcactccctggcctttcgacatcctgcttccatgttcatgttcttggcgttgacgagcaatg 780
Db 721 tgcttcactccctggcctttcgacatcctgcttccatgttcatgttcttggcgttgacgagcaatg 780
QY 781 catctggctgaagaatgcacagatgctgctcgtacgttaccacaaagcagtcagtcagtc 840
Db 781 catctggctgaagaatgcacagatgctgctcgtacgttaccacaaagcagtcagtcagtc 840
QY 841 atcataattggccttgcacccgcttcccatatacaaatcgcagttctgtatggaattaca 900
Db 841 atcataattggccttgcacccgcttcccatatacaaatcgcagttctgtatggaattaca 900
QY 901 gatctcgactctattctaaagtcccgccggtctctatattccattcgcagacaatgacgcagtc 960
Db 901 gatctcgactctattctaaagtcccgccggtctctatattccattcgcagacaatgacgcagtc 960
QY 961 ctctcggctcgtcagttttgcaacgggtcctctcctcgtacgttgcgtggtggtcgttcttctt 1020
Db 961 ctctcggctcgtcagttttgcaacgggtcctctcctcgtacgttgcgtggtggtcgttcttctt 1020
QY 1021 gccctcaacgctgtacaagagactgcgtctcgcactcactcactcactcactcactcactcact 1080
Db 1021 gccctcaacgctgtacaagagactgcgtctcgcactcactcactcactcactcactcactcact 1080
QY 1081 gggctgggtattttccactcatctcgaacgcattcactcactcactcactcactcactcactcact 1140
Db 1081 gggctgggtattttccactcatctcgaacgcattcactcactcactcactcactcactcactcact 1140
QY 1141 tctctattcgcgacctggggaattcttgccacatcgcggtatgtatatatttctaggttcttagc 1200
Db 1141 tctctattcgcgacctggggaattcttgccacatcgcggtatgtatatatttctaggttcttagc 1200
QY 1201 acagctttcaatgccttgggtcaattccgcgcgttcttactccagcaactccttcttcttctgac 1260
Db 1201 acagctttcaatgccttgggtcaattccgcgcgttcttactccagcaactccttcttcttctgac 1260
QY 1261 ccaatcgccctactcctctacccaaagcgagatccaaagtcttctgcccagacactcgtgct 1320
Db 1261 ccaatcgccctactcctctacccaaagcgagatccaaagtcttctgcccagacactcgtgct 1320
QY 1321 ttgtgttaccgctgggaatcgggtttcttctgttgcgtatgtgcgtggtgttcttcttcttcttct 1380
Db 1321 ttgtgttaccgctgggaatcgggtttcttctgttgcgtatgtgcgtggtgttcttcttcttcttct 1380
QY 1381 gtcaccactgtgttttccagcttccactgacgtgcctacggcgcgtgcacccatgaat 1440
Db 1381 gtcaccactgtgttttccagcttccactgacgtgcctacggcgcgtgcacccatgaat 1440
QY 1441 tacacaagtgcgattatagcggttgcaacttgcgttgcgttgcgttgcgttgcgttgcgttgcgt 1500
Db 1441 tacacaagtgcgattatagcggttgcaacttgcgttgcgttgcgttgcgttgcgttgcgttgcgt 1500


```
RESULT 5
US-09-882-694-7
; Sequence 7, Application US/09882694
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; FILE REFERENCE: 35718/208255
; CURRENT APPLICATION NUMBER: US/09/882,694
; CURRENT FILING DATE: 2001-06-12
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1578
; TYPE: DNA
; ORGANISM: Exophiala spinifera
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)..(0)
; OTHER INFORMATION: permease, fully spliced cDNA
US-09-882-694-7
```

```
Query Match      100.0%; Score 1578; DB 33; Length 1578;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1578; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggactccagaccagtgatagcggcgagagaagggcgagacaagggcgagacaacgaagaac 60
Db 1 atggactccagaccagtgatagcggcgagagaagggcgagacaagggcgagacaacgaagaac 60

QY 61 acagagacggcgggcgagtggtgctgcgtccgagtcacctgaacgttccctctggagaagaaa 120
Db 61 acagagacggcgggcgagtggtgctgcgtccgagtcacctgaacgttccctctggagaagaaa 120

QY 121 caatttggcaccatcaccatcgtgtccttggcccttggattgtgatttgcaacagttggctggt 180
Db 121 caatttggcaccatcaccatcgtgtccttggcccttggattgtgatttgcaacagttggctggt 180

QY 181 atctcaggcagtcctccagctcgcctactagcggggggggcccgctcactctcctttacggc 240
Db 181 atctcaggcagtcctccagctcgcctactagcggggggggcccgctcactctcctttacggc 240

QY 241 atcctaatacagctactctcgtctacatctgcactcgttctcattagccgaactgaccagc 300
Db 241 atcctaatacagctactctcgtctacatctgcactcgttctcattagccgaactgaccagc 300

QY 301 gtctaccgagctgcgggtggccaatatcatcttgcgtcgtatcctggcaccacaaatcaatc 360
Db 301 gtctaccgagctgcgggtggccaatatcatcttgcgtcgtatcctggcaccacaaatcaatc 360

QY 361 aatcgagcatttcatacgtgtgcggactcgtgtcgttgcgttgcattgcgtatcgatcgga 420
Db 361 aatcgagcatttcatacgtgtgcggactcgtgtcgttgcgttgcattgcgtatcgatcgga 420

QY 421 agctcagtgaccatgatacctgctcaacagatcccgcgctgatagcgcctatagtcac 480
Db 421 agctcagtgaccatgatacctgctcaacagatcccgcgctgatagcgcctatagtcac 480

QY 481 acatactcccaggattcgtggcatgtcttctcctcatctacgaggggagtcgcgtggtggtg 540
Db 481 acatactcccaggattcgtggcatgtcttctcctcatctacgaggggagtcgcgtggtggtg 540

QY 541 ctctgttccaactgtttgcccctgaaagaaaccccttgggttccatgaaatcggattcggc 600
Db 541 ctctgttccaactgtttgcccctgaaagaaaccccttgggttccatgaaatcggattcggc 600
```

```
QY 601 ctacgatacgtctctctcgtgatctctcttcttatcgcacattcttagcgcgggtccaaacccaag 660
Db 601 ctacgatacgtctctctcgtgatctctcttcttatcgcacattcttagcgcgggtccaaacccaag 660

QY 661 gctccaaactcacaaggatggaactgcttgaggacaactatactaggctgggtccgacggcgctc 720
Db 661 gctccaaactcacaaggatggaactgcttgaggacaactatactaggctgggtccgacggcgctc 720

QY 721 tgccttcactcctgggccccttctcgacatcctcgtcttcatgttcatgttgcgttgacgcagcaatg 780
Db 721 tgccttcactcctgggccccttctcgacatcctcgtcttcatgttcatgttgcgttgacgcagcaatg 780

QY 781 catctggctgaagaatgcacagatgctgctcgtcgtacggtaccacaaagcagtggtcagtgca 840
Db 781 catctggctgaagaatgcacagatgctgctcgtcgtacggtaccacaaagcagtggtcagtgca 840

QY 841 atcataattggcttctgcaccgccttctccatatatacaatcgcaggtctctgtatggaattaca 900
Db 841 atcataattggcttctgcaccgccttctccatatatacaatcgcaggtctctgtatggaattaca 900

QY 901 gatctcgactctattcttaagtctccgcggcgtatatattccattcgcagacaatgacgcagtc 960
Db 901 gatctcgactctattcttaagtctccgcggcgtatatattccattcgcagacaatgacgcagtc 960

QY 961 ctctcggctcgtcagtttgcacaagctcctctcctcatgtggcggtatcgtgatggcctctcttc 1020
Db 961 ctctcggctcgtcagtttgcacaagctcctctcctcatgtggcggtatcgtgatggcctctcttc 1020

QY 1021 gccctcaacgctgtacaagagactcgtctctcgactcactcactggagcttctgcccggggacaat 1080
Db 1021 gccctcaacgctgtacaagagactcgtctctcgactcactcactggagcttctgcccggggacaat 1080

QY 1081 gggctggtatttcccaactcatctcgaaacgcattcatcccgctggcgaagtctcctgttgg 1140
Db 1081 gggctggtatttcccaactcatctcgaaacgcattcatcccgctggcgaagtctcctgttgg 1140

QY 1141 tctctattcgcgacctggggaattctcggccacatgcggatgtatatattctagggttcttagc 1200
Db 1141 tctctattcgcgacctggggaattctcggccacatgcggatgtatatattctagggttcttagc 1200

QY 1201 acagctttcaatgccttgggtcaattcgcgcgctgtgactccagcaactctccttctcctgac 1260
Db 1201 acagctttcaatgccttgggtcaattcgcgcgctgtgactccagcaactctccttctcctgac 1260

QY 1261 ccaatcgccctactcctctaccacaaagcgagatccaaagtctcctggcgcgactcgtgct 1320
Db 1261 ccaatcgccctactcctctaccacaaagcgagatccaaagtctcctggcgcgactcgtgct 1320

QY 1321 tttgtgtaccgcgtgggaatcgggttcttggtcgaatgtgctagcgggtggtcttcaacgtcc 1380
Db 1321 tttgtgtaccgcgtgggaatcgggttcttggtcgaatgtgctagcgggtggtcttcaacgtcc 1380

QY 1381 gtcaccactgtgttttctcagcttcccaactgacctgacctgacctgacctgacctgacctgacct 1440
Db 1381 gtcaccactgtgttttctcagcttcccaactgacctgacctgacctgacctgacctgacctgacct 1440

QY 1441 tacacaagtgcgattatagggcgttgcaacttgctcttgggtcttgaactgggtcgtgcat 1500
Db 1441 tacacaagtgcgattatagggcgttgcaacttgctcttgggtcttgaactgggtcgtgcat 1500

QY 1501 gccagggaagcattatcaggggaccccaacttggagcttgacgggacgggtcgtcgggagcagaa 1560
Db 1501 gccagggaagcattatcaggggaccccaacttggagcttgacgggacgggtcgtcgggagcagaa 1560

QY 1561 tttcaagttggggcccatga 1578
Db 1561 tttcaagttggggcccatga 1578
```

```
RESULT 6
US-09-882-694A-7
; Sequence 7, Application US/09882694A
```



```

? GENERAL INFORMATION:
? APPLICANT: Duwick, Jon
? APPLICANT: Maddox, Joyce
? APPLICANT: Gilliam, Jacob
? APPLICANT: Folkerts, Otto
? APPLICANT: Crasta, Oswald R.
? TITLE OF INVENTION: Compositions and Methods for Fumonisin
? TITLE OF INVENTION: Detoxification
? FILE REFERENCE: 5718-111
? CURRENT APPLICATION NUMBER: US/09/882,694A
? CURRENT FILING DATE: 2001-06-15
? PRIOR APPLICATION NUMBER: 09/351,224
? PRIOR FILING DATE: 1999-07-12
? NUMBER OF SEQ ID NOS: 11
? SOFTWARE: FastSEQ for Windows Version 4.0
? SEQ ID NO 7
? LENGTH: 1578
? TYPE: DNA
? ORGANISM: Exophiala spinifera
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (0)...(0)
? OTHER INFORMATION: permease, fully spliced cDNA
US-09-882-694A-7

```

		Query Match	100.0%;	Score 1578;	DB 33;	Length 1578;
		Best Local Similarity	100.0%;	Pred. NO. 0;		
		Matches 1578;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	atggactccagacc	aagtggatac	gcgagaaa	aggcgggaca	aggcagacaacgaagaac 60
Db	1	atggactccagacc	aagtggatac	gcgagaaa	aggcgggaca	aggcagacaacgaagaac 60
QY	61	acagagacggcg	ggcgaggtg	cgctccgagtc	ccctgaacgtt	ccctctgagagaagaaa 120
Db	61	acagagacggcg	ggcgaggtg	cgctccgagtc	ccctgaacgtt	ccctctgagagaagaaa 120
QY	121	caatttggcacc	atcaccatc	gtgtcctt	ggccctt	gtgatttgcacacagttgggctggt 180
Db	121	caatttggcacc	atcaccatc	gtgtcctt	ggccctt	gtgatttgcacacagttgggctggt 180
QY	181	atctcaggcagt	ctccagctc	gcgcctact	agcggggggcc	cgtcactctccttttacggc 240
Db	181	atctcaggcagt	ctccagctc	gcgcctact	agcggggggcc	cgtcactctccttttacggc 240
QY	241	atcctaatac	agttactct	cgctctacat	ctgcctt	cttctcattagccgaactgaccagc 300
Db	241	atcctaatac	agttactct	cgctctacat	ctgcctt	cttctcattagccgaactgaccagc 300
QY	301	gtctaccgact	gcgggtgg	cccaatat	catcttt	gctgcgacccctggcaccacaaatcaatc 360
Db	301	gtctaccgact	gcgggtgg	cccaatat	catcttt	gctgcgacccctggcaccacaaatcaatc 360
QY	361	aatcggagatt	tcatacgt	gtgcggact	cgtgtcg	tgttctcatggatcgctatcgga 420
Db	361	aatcggagatt	tcatacgt	gtgcggact	cgtgtcg	tgttctcatggatcgctatcgga 420
QY	421	agctcagtgac	catgatac	ctgtctcaac	agatccccg	gcgctgatagcgcctatagtcac 480
Db	421	agctcagtgac	catgatac	ctgtctcaac	agatccccg	gcgctgatagcgcctatagtcac 480
QY	481	acatactccag	gattcgt	gtggcatgt	cttctcct	catctacagaggagtcgcgtcggtggtg 540
Db	481	acatactccag	gattcgt	gtggcatgt	cttctcct	catctacagaggagtcgcgtcggtggtg 540
QY	541	ctcttgttca	acttgtt	tgccctgaaa	agaaaccc	cttgggttcatgaaatcgattcggc 600
Db	541	ctcttgttca	acttgtt	tgccctgaaa	agaaaccc	cttgggttcatgaaatcgattcggc 600
QY	601	ctcacgatac	gtctctct	gtgatctct	cttattat	ogccattctagcgcggtccaaaccccaag 660
Db	601	ctcacgatac	gtctctct	gtgatctct	cttattat	ogccattctagcgcggtccaaaccccaag 660

QY	661	gctccaaactcacaggtatggactgcttggagcaactalactggctggtccgacggcgctc	720
Db	661		720
QY	721	tgcttcacctcctgggccccttcgacacatccctgcttcacgttcatgttccattggccttggacgcgacgaatg	780
Db	721		780
QY	781	catctggctgaagaatgcacagatgctgctcgtagctacccaagcagtggtcagtgca	840
Db	781		840
QY	841	atcataattggcttctgcacgcgcctttccatatatacaaatgcgcagttctgtatgggaattaca	900
Db	841		900
QY	901	gattctgactctattcttaagtctccgcgggctctatattccattccgagacaaatgacgcagtcct	960
Db	901		960
QY	961	cttcgggtcgctcagttttgcacacgggtccctctcatgtggcgggtatcgtgatggccttctctc	1020
Db	961		1020
QY	1021	gccctcaacgcctgtacaagagactgcgtctcgactcacctggagccttgcgcgggacaaat	1080
Db	1021		1080
QY	1081	gggctgggtattttccactcatctctgaacgcattcatcccccgtggcaagttcctgttttg	1140
Db	1081		1140
QY	1141	tctctattcgcgacctggggaattcttgccacatgcggatgtatatattctaggttcttagc	1200
Db	1141		1200
QY	1201	acagctttcaatgccttggccaattccgcgcttgactccagcaactctccttccctgactc	1260
Db	1201		1260
QY	1261	ccaatgcacctactcctctaccacaaagcgagatccaaaagtctctgcgcgacactcgtgct	1320
Db	1261		1320
QY	1321	tttggttaaccgcgtggaaatcggtttctggtcgaatgtgctagcgggtggtcttcaacgtcc	1380
Db	1321		1380
QY	1381	gtcaccaactggtttttcaagcttcccactgaccggtccctacgcgcgtcgaacccatgaat	1440
Db	1381		1440
QY	1441	tacacaagtgcattataggcgttgcaacttgctcttggtgtcttgaaactgggtcggtgcatt	1500
Db	1441		1500
QY	1501	gccaggaaagcattatcagggaaccccacttggaagcttgacggacgggtcgtcggagcagaa	1560
Db	1501		1560
QY	1561	tttcaagtggggccatga	1578
Db	1561		1578

```

RESULT 7
US-09-351-224E-6
; Sequence 6, Application US/09351224E
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob

```


; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; FILE REFERENCE: 5718-111
; CURRENT APPLICATION NUMBER: US/09/351,224E
; CURRENT FILING DATE: 1999-07-12
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 1764
; TYPE: DNA
; ORGANISM: Exophiala spinifera
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: permease, partially spliced cDNA
US-09-351-224E-6

Query Match 100.0%; Score 1578; DB 17; Length 1764;
Best Local Similarity 100.0%; Pred. NO. 0;
Matches 1578; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	atggactccagacc	aaagtggata	cgcgagaa	aggcgga	aaaggcagaca	acgaagaac	60					
Db	5	atggactccagacc	aaagtggata	cgcgagaa	aggcgga	aaaggcagaca	acgaagaac	64					
QY	61	acagagacggcg	ggcagggt	ggcgcc	gagtc	cccgaa	cgttcctctg	gagaagaaa	120				
Db	65	acagagacggcg	ggcagggt	ggcgcc	gagtc	cccgaa	cgttcctctg	gagaagaaa	124				
QY	121	caatttggcacc	atcaccat	cgtgtc	cttggc	ctttgg	catttgc	aacagttgg	gctggt	180			
Db	125	caatttggcacc	atcaccat	cgtgtc	cttggc	ctttgg	catttgc	aacagttgg	gctggt	184			
QY	181	atctcaggcag	tctccag	ctcgcc	tactag	cggggg	ggcccg	tcaact	tctcttac	ggc	240		
Db	185	atctcaggcag	tctccag	ctcgcc	tactag	cggggg	ggcccg	tcaact	tctcttac	ggc	244		
QY	241	atcctaatac	gtactctc	gtctac	atctgc	atcgtctt	tcattag	cgcga	aaactg	accagc	300		
Db	245	atcctaatac	gtactctc	gtctac	atctgc	atcgtctt	tcattag	cgcga	aaactg	accagc	304		
QY	301	gtctaccggac	tgcgggtg	ggccaata	tcatctt	tgctg	atccttg	gcac	caaaaat	caatc	360		
Db	305	gtctaccggac	tgcgggtg	ggccaata	tcatctt	tgctg	atccttg	gcac	caaaaat	caatc	364		
QY	361	aatcggagca	tttcaat	cgtgtg	cggaact	cgctg	ctgtt	gcatt	gcgtat	cgcga	420		
Db	365	aatcggagca	tttcaat	cgtgtg	cggaact	cgctg	ctgtt	gcatt	gcgtat	cgcga	424		
QY	421	agctcagtg	acatgat	atcctg	ctcaac	agatcccc	ggcgctg	atagcc	gcctat	agtccac	480		
Db	425	agctcagtg	acatgat	atcctg	ctcaac	agatcccc	ggcgctg	atagcc	gcctat	agtccac	484		
QY	481	acatactccc	aggattcg	tggcat	gtctt	ctctcat	cagagg	gagtc	gcgctg	ggtg	540		
Db	485	acatactccc	aggattcg	tggcat	gtctt	ctctcat	cagagg	gagtc	gcgctg	ggtg	544		
QY	541	ctcttgttca	actgtt	tgccctg	aaaaa	gaaac	cccttgg	gttccat	gaaatc	ggattc	ggc	600	
Db	545	ctcttgttca	actgtt	tgccctg	aaaaa	gaaac	cccttgg	gttccat	gaaatc	ggattc	ggc	604	
QY	601	ctcacgatc	gctctct	ctctg	tgatct	ctctt	tatcg	ccattc	tagcgc	gggtcc	aaaccc	aaag	660
Db	605	ctcacgatc	gctctct	ctctg	tgatct	ctctt	tatcg	ccattc	tagcgc	gggtcc	aaaccc	aaag	664
QY	661	gtccaaactc	acaggtat	ggactg	cttgg	agcaact	atactg	gctg	gttcc	gacg	gcgtc	720	
Db	665	gtccaaactc	acaggtat	ggactg	cttgg	agcaact	atactg	gctg	gttcc	gacg	gcgtc	724	
QY	721	tgcttcatc	cttgggc	cttctc	gacat	cctg	cttcat	gttcat	tggctt	ggacg	cagcaatg	780	

Db	725	tgcttcatc	cttgggc	cttctc	gacat	ctcgt	cttc	atgtt	cttc	atgtt	ggcgt	ggacg	caaatg	784			
QY	781	catctggct	gaagaat	gcacag	atgct	gctg	tac	ggttac	ccca	aaagc	agtggt	ctc	agtga	840			
Db	785	catctggct	gaagaat	gcacag	atgct	gctg	tac	ggttac	ccca	aaagc	agtggt	ctc	agtga	844			
QY	841	atcataa	attggct	cttctg	caccg	ctttcc	atata	caaat	cgc	agttct	gtat	gga	attaca	900			
Db	845	atcataa	attggct	cttctg	caccg	ctttcc	atata	caaat	cgc	agttct	gtat	gga	attaca	904			
QY	901	gatctgac	tcttat	ctta	agttcc	gcgg	ctatat	attcc	atcg	agaca	aatgac	gcag	tcc	960			
Db	905	gatctgac	tcttat	ctta	agttcc	gcgg	ctatat	attcc	atcg	agaca	aatgac	gcag	tcc	964			
QY	961	cttcgg	tcgctc	agttt	gcaac	gggtcc	ctctc	atg	tgg	cgggtat	cgtg	atg	ggc	cttcttc	1020		
Db	965	cttcgg	tcgctc	agttt	gcaac	gggtcc	ctctc	atg	tgg	cgggtat	cgtg	atg	ggc	cttcttc	1024		
QY	1021	gcctca	acgctg	taca	agag	actgc	gtctc	gact	cac	ctgg	agctt	tgcc	ggg	acaat	1080		
Db	1025	gcctca	acgctg	taca	agag	actgc	gtctc	gact	cac	ctgg	agctt	tgcc	ggg	acaat	1084		
QY	1081	gggctg	gtattt	ttcc	actcat	ctc	gaac	gcat	cccc	gcgt	gca	agttc	ctg	tttgg	1140		
Db	1085	gggctg	gtattt	ttcc	actcat	ctc	gaac	gcat	cccc	gcgt	gca	agttc	ctg	tttgg	1144		
QY	1141	ttctatt	cgcg	ac	ctgg	ggaatt	ctt	ctg	gcac	atgc	ggtat	ttct	atg	ttctag	1200		
Db	1145	ttctatt	cgcg	ac	ctgg	ggaatt	ctt	ctg	gcac	atgc	ggtat	ttct	atg	ttctag	1204		
QY	1201	acagctt	ccaat	gcctt	gg	tc	caatt	ccgc	gctg	tact	ccag	caact	ctc	ctctg	1260		
Db	1205	acagctt	ccaat	gcctt	gg	tc	caatt	ccgc	gctg	tact	ccag	caact	ctc	ctctg	1264		
QY	1261	ccaat	cgcct	actc	ctct	tac	caaaa	agc	gagat	ccaaa	ggtt	ctt	g	ccgag	1320		
Db	1265	ccaat	cgcct	actc	ctct	tac	caaaa	agc	gagat	ccaaa	ggtt	ctt	g	ccgag	1324		
QY	1321	ttt	gtgt	tac	ccg	ctg	ggaat	cgg	gttt	ctt	cggt	caat	gt	g	ctag	1380	
Db	1325	ttt	gtgt	tac	ccg	ctg	ggaat	cgg	gttt	ctt	cggt	caat	gt	g	ctag	1384	
QY	1381	gtc	acc	actg	tgt	gttt	ttc	agctt	cccc	actg	acc	gtg	cc	gcgc	gtca	1440	
Db	1385	gtc	acc	actg	tgt	gttt	ttc	agctt	cccc	actg	acc	gtg	cc	gcgc	gtca	1444	
QY	1441	tac	aca	agt	gcg	attat	agc	gtt	gc	actt	gct	ctt	g	gtg	ctt	g	1500
Db	1445	tac	aca	agt	gcg	attat	agc	gtt	gc	actt	gct	ctt	g	gtg	ctt	g	1504
QY	1501	gcc	agg	aag	catt	atc	agg	gac	cccc	actt	gg	agctt	g	acg	g	g	1560
Db	1505	gcc	agg	aag	catt	atc	agg	gac	cccc	actt	gg	agctt	g	acg	g	g	1564
QY	1561	ttt	caa	gtt	ggg	ccat	ga	1578									
Db	1565	ttt	caa	gtt	ggg	ccat	ga	1582									

RESULT 8
US-09-677-488A-6
; Sequence 6, Application US/09677488A
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; TITLE OF INVENTION: Detoxification
; FILE REFERENCE: 35718/204100
; CURRENT APPLICATION NUMBER: US/09/677,488A

```
; CURRENT FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 1764
; TYPE: DNA
; ORGANISM: Exophiala spinifera
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: permease, partially spliced cDNA
US-09-677-488A-6

Query Match      100.0%; Score 1578; DB 26; Length 1764;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1578; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggactccagaccagtgatagcggcgagaaagcgcgagacagcagacagaaac 60
Db 1 |||||||
QY 5 atggactccagaccagtgatagcggcgagaaagcgcgagacagcagacagaaac 64
Db 5 |||||||
QY 61 acagagacggcgggcgaggtggtgcgtccaggtccctgaacgttctctggagaagaa 120
Db 61 |||||||
QY 65 acagagacggcgggcgaggtggtgcgtccaggtccctgaacgttctctggagaagaa 124
Db 65 |||||||
QY 121 caattggcaccatcacatcgtgtccttggcctttgtgatttgaacagttggctggt 180
Db 121 |||||||
QY 125 caattggcaccatcacatcgtgtccttggcctttgtgatttgaacagttggctggt 184
Db 125 |||||||
QY 181 atctcaggcagtcctccagtcgcccactactagcggggggggcccgctcactctcttaacgc 240
Db 181 |||||||
QY 185 atctcaggcagtcctccagtcgcccactactagcggggggggcccgctcactctcttaacgc 244
Db 185 |||||||
QY 241 atctaatcagtaactctctacatctgcactgcgtcttctcattagcgaactgaccagc 300
Db 241 |||||||
QY 245 atctaatcagtaactctctacatctgcactgcgtcttctcattagcgaactgaccagc 304
Db 245 |||||||
QY 301 gtctaccgactgcccgtggcgaatatcatcttgcgtcgatccctggcaccacaaatcaatc 360
Db 301 |||||||
QY 305 gtctaccgactgcccgtggcgaatatcatcttgcgtcgatccctggcaccacaaatcaatc 364
Db 305 |||||||
QY 361 aatcgagacatttcatactggtgaggactggtgcttgcgtcttctcattgactgctatcgga 420
Db 361 |||||||
QY 365 aatcgagacatttcatactggtgaggactggtgcttgcgtcttctcattgactgctatcgga 424
Db 365 |||||||
QY 421 agctcagtgaccatgatacctgctcaacagatcccgcgctgactgactgactgactgactgact 480
Db 421 |||||||
QY 425 agctcagtgaccatgatacctgctcaacagatcccgcgctgactgactgactgactgactgact 484
Db 425 |||||||
QY 481 acatactcccaggattcgtggcagatgtcttctcctcattctacagggagtgctggtggtg 540
Db 481 |||||||
QY 485 acatactcccaggattcgtggcagatgtcttctcctcattctacagggagtgctggtggtg 544
Db 485 |||||||
QY 541 ctcttgttcaactgttttgcctgaaagaaaccccttgggttcatgaaatcggattcggc 600
Db 541 |||||||
QY 545 ctcttgttcaactgttttgcctgaaagaaaccccttgggttcatgaaatcggattcggc 604
Db 545 |||||||
QY 601 ctcaagatcgtctctctctgactgctcttcttgcgacttctagcgcggttcaaccccaag 660
Db 601 |||||||
QY 605 ctcaagatcgtctctctctgactgctcttcttgcgacttctagcgcggttcaaccccaag 664
Db 605 |||||||
QY 661 gctccaaactcacaggtatgactgcttggagcaactatactggtggtgcgacgctc 720
Db 661 |||||||
QY 665 gctccaaactcacaggtatgactgcttggagcaactatactggtggtgcgacgctc 724
Db 665 |||||||
QY 721 tgccttcactcctggccttttcacatcctgcttctcattgttctggtgagcagcaatg 780
Db 721 |||||||
QY 725 tgccttcactcctggccttttcacatcctgcttctcattgttctggtgagcagcaatg 784
Db 725 |||||||
QY 781 catctggtgaagaatgcacagatgctgctcgtacggtacccaaagcagtggtcagtgca 840
Db 781 |||||||
```

```
Db 785 catctggctgaagaatgcacagatgctgctcgtacggtacccaaagcagtggtcagtgca 844
QY 841 atcataaattggtcttctgcaccgcttccatatacaaatcgaggtctgtatggaattaca 900
Db 845 atcataaattggtcttctgcaccgcttccatatacaaatcgaggtctgtatggaattaca 904
QY 901 gatctcgactctattctaaagtcccgcggtctatatattccattcgagacaatgacgcagtc 960
Db 905 gatctcgactctattctaaagtcccgcggtctatatattccattcgagacaatgacgcagtc 964
QY 961 ctctcggtcgtcagttttgcaacggtcctctcatgtggcggtatcgtgatgaccttcttc 1020
Db 965 ctctcggtcgtcagttttgcaacggtcctctcatgtggcggtatcgtgatgaccttcttc 1024
QY 1021 gccctcaacgctgtacaagagactgctcgtcgtcgtcgtcgttggcggttggcggttgg 1080
Db 1025 gccctcaacgctgtacaagagactgctcgtcgtcgtcgtcgttggcggttggcggttgg 1084
QY 1081 gggctggtattttccactcatctcgaacgcatctcatcccgctggcgaagtccctgtttgg 1140
Db 1085 gggctggtattttccactcatctcgaacgcatctcatcccgctggcgaagtccctgtttgg 1144
QY 1141 tctctattcgcgacctggggaattctctggccacatgcggatgtatatatttctaggttctagc 1200
Db 1145 tctctattcgcgacctggggaattctctggccacatgcggatgtatatatttctaggttctagc 1204
QY 1201 acagctttcaatgcttgggtcaattcccgcttggactccagcaactctcttctcgtgac 1260
Db 1205 acagctttcaatgcttgggtcaattcccgcttggactccagcaactctcttctcgtgac 1264
QY 1261 ccaatcgccctactcctctacccaaagcgagatcccaagtcttcttgcgagcactcgtgct 1320
Db 1265 ccaatcgccctactcctctacccaaagcgagatcccaagtcttcttgcgagcactcgtgct 1324
QY 1321 tttgtgttaccgctggaatcggtttcttgggtcaatgtgctagcgttgggtcttcaacgtcc 1380
Db 1325 tttgtgttaccgctggaatcggtttcttgggtcaatgtgctagcgttgggtcttcaacgtcc 1384
QY 1381 gtcaccactgtgtttttcagcttcccaactgacgttgcctacggtccggtcaacccatgaat 1440
Db 1385 gtcaccactgtgtttttcagcttcccaactgacgttgcctacggtccggtcaacccatgaat 1444
QY 1441 tacacaagtgcgattataggcgttgcacttgccttgggtgcttgaactgggtcgtgcat 1500
Db 1445 tacacaagtgcgattataggcgttgcacttgccttgggtgcttgaactgggtcgtgcat 1504
QY 1501 gccaggaagcattatcagggagcccccacttggagcttgacggaggttgcggagcagaa 1560
Db 1505 gccaggaagcattatcagggagcccccacttggagcttgacggaggttgcggagcagaa 1564
QY 1561 tttcaagtgtggccatga 1578
Db 1565 tttcaagtgtggccatga 1582
```

```
RESULT 9
US-09-677-682A-6
; Sequence 6, Application US/09677682A
; GENERAL INFORMATION:
; APPLICANT: DuVick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; TITLE OF INVENTION: Detoxification
; FILE REFERENCE: 35718/204101
; CURRENT APPLICATION NUMBER: US/09/677,682A
; CURRENT FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
```


OTHER INFORMATION: permease, partially spliced cDNA
US-09-882-694-6

Query Match 100.0%; Score 1578; DB 33; Length 1764;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1578; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggactcagacacaaagtggatagcgcgagaaagcggagacaaagcagacaaagaaac 60
Db 5 atggactcagacacaaagtggatagcgcgagaaagcggagacaaagcagacaaagaaac 64
QY 61 acagagacggcgcgaggggtggtgcgtccgagtcctcctgaacgttccctctggagagaa 120
Db 65 acagagacggcgcgaggggtggtgcgtccgagtcctcctgaacgttccctctggagagaa 124
QY 121 caatttggcaccatcaccaatcgtgtccttggccttctgtgatttgcacagttgggtggt 180
Db 125 caatttggcaccatcaccaatcgtgtccttggccttctgtgatttgcacagttgggtggt 184
QY 181 atctcaggcagtcctccagtcgcctactagcggggggggcccgctcactctcctttacggc 240
Db 185 atctcaggcagtcctccagtcgcctactagcggggggggcccgctcactctcctttacggc 244
QY 241 atcctaatacgtactctcgtctacatctgcacgtcgttctcattagccgaactgaccagc 300
Db 245 atcctaatacgtactctcgtctacatctgcacgtcgttctcattagccgaactgaccagc 304
QY 301 gtctaccgactgcccgtggcccaataatcttttgcgtcgtatcctcggcaccacaaatcaatc 360
Db 305 gtctaccgactgcccgtggcccaataatcttttgcgtcgtatcctcggcaccacaaatcaatc 364
QY 361 aatcggagcatttcatacagtcgtgcggactcgttgcgttgcgttctcattggtatcgga 420
Db 365 aatcggagcatttcatacagtcgtgcggactcgttgcgttgcgttctcattggtatcgga 424
QY 421 agctcagtgaccatgatacctgctcaacagatcccgccgctgtagccgcctatagtcac 480
Db 425 agctcagtgaccatgatacctgctcaacagatcccgccgctgtagccgcctatagtcac 484
QY 481 acatactccaggattcgtggcagtcgttctcctcactctacgagggagtcgctggtggtg 540
Db 485 acatactccaggattcgtggcagtcgttctcctcactctacgagggagtcgctggtggtg 544
QY 541 ctcttgttcaactgtttgcccctgaaagaaaccccttgggttccatgaaatcggattcggc 600
Db 545 ctcttgttcaactgtttgcccctgaaagaaaccccttgggttccatgaaatcggattcggc 604
QY 601 ctcacgagtcgtctcttctgagtcctcttctcctcattctagcgggtcccaaccccaag 660
Db 605 ctcacgagtcgtctcttctgagtcctcttctcctcattctagcgggtcccaaccccaag 664
QY 661 gctccaaactcacaggatgagtcgttggagcaactatactggtggtccgacggcgtc 720
Db 665 gctccaaactcacaggatgagtcgttggagcaactatactggtggtccgacggcgtc 724
QY 721 tgcctcactcctgggcccctttcgacatcctgcttcatgttcatgttgcgttggacgcagcaatg 780
Db 725 tgcctcactcctgggcccctttcgacatcctgcttcatgttcatgttgcgttggacgcagcaatg 784
QY 781 catctggtgaagaatgcacagatgctgctcagcagtcaccccaagcagtcagtcagtcga 840
Db 785 catctggtgaagaatgcacagatgctgctcagcagtcaccccaagcagtcagtcagtcga 844
QY 841 atcataatggcttctgcacccgcttccatatacaaatcgcagtcgttatggaattaca 900
Db 845 atcataatggcttctgcacccgcttccatatacaaatcgcagtcgttatggaattaca 904
QY 901 gatctcagactctattctaaagtcccgccggtatattccattcggagacaaatgacgcagtc 960
Db 905 gatctcagactctattctaaagtcccgccggtatattccattcggagacaaatgacgcagtc 964
QY 961 ctctcgtcgtcagtttttgcacacgggtcctctcactgtggtggtggtatggtggtccttctc 1020

Db 965 ctctcgtcgtcagtttttgcacacgggtcctctcactgttggcgtatcgtgagtccttctctc 1024
QY 1021 gccctcaacgctgtacaagagactgcgtctcagactcaccttggagccttcttgcgggagacaat 1080
Db 1025 gccctcaacgctgtacaagagactgcgtctcagactcaccttggagccttcttgcgggagacaat 1084
QY 1081 gggctggtattttccactcatctcgaacgcatctcctcccgctggcaagtctcctgttgg 1140
Db 1085 gggctggtattttccactcatctcgaacgcatctcctcccgctggcaagtctcctgttgg 1144
QY 1141 tctctattcgcgacctggggaattcttggccacatgcggatgtatatcttctaggttctagc 1200
Db 1145 tctctattcgcgacctggggaattcttggccacatgcggatgtatatcttctaggttctagc 1204
QY 1201 acagctttcaatgccttgggtcaattcccgcttgtactccacgaactctccttctcagtc 1260
Db 1205 acagctttcaatgccttgggtcaattcccgcttgtactccacgaactctccttctcagtc 1264
QY 1261 ccaatgcctcactcctctacccaaagcggagatccaaagtcttctgcccagacactcgtgct 1320
Db 1265 ccaatgcctcactcctctacccaaagcggagatccaaagtcttctgcccagacactcgtgct 1324
QY 1321 ttctgtttaccgctgggaatcgggtttcttgggtcaatgtgctagcgggtggtcttctcagtc 1380
Db 1325 ttctgtttaccgctgggaatcgggtttcttgggtcaatgtgctagcgggtggtcttctcagtc 1384
QY 1381 gtcaccactgtgtttctcagcttcccaactgacccgtgcctacggccgctcaacccatgaat 1440
Db 1385 gtcaccactgtgtttctcagcttcccaactgacccgtgcctacggccgctcaacccatgaat 1444
QY 1441 tacacaagtgcgattatagcgttgcacttgccttgggtgttgaactgggtcgtgcat 1500
Db 1445 tacacaagtgcgattatagcgttgcacttgccttgggtgttgaactgggtcgtgcat 1504
QY 1501 gccaggaagcattatcagggaccccaacttggagcttgacggagcgggtcgtcggagagaa 1560
Db 1505 gccaggaagcattatcagggaccccaacttggagcttgacggagcgggtcgtcggagagaa 1564
QY 1561 ttctcaagttgggcccata 1578
Db 1565 ttctcaagttgggcccata 1582

RESULT 12
US-09-882-694A-6
; Sequence 6, Application US/09882694A
; GENERAL INFORMATION:
; APPLICANT: Duivick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; TITLE OF INVENTION: Detoxification
; FILE REFERENCE: 5718-111
; CURRENT APPLICATION NUMBER: US/09/882,694A
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 1764
; TYPE: DNA
; ORGANISM: Exophiala spinifera
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: permease, partially spliced cDNA
US-09-882-694A-6

Db 150 ggccttctgattgcaacagttgggtatctcaggcagtcctccagctcgcctact 209
Qy 210 agcggggggcccgctcaactctctttacggcatcctaatacagtaactctcgtctacatctg 269
Db 210 agcggggggcccgctcaactctctttacggcatcctaatacagtaactctcgtctacatctg 269
Qy 270 catcgccttctcattagccgaactgaccagcgtctacccgactgccggtggccaatatca 329
Db 270 catcgccttctcattagccgaactgaccagcgtctacccgactgccggtggccaatatca 329
Qy 330 ttttgctcgtatccctggcaccacaaatacaatcaatcgaatcggagcatttcatacgtgtgaggact 389
Db 330 ttttgctcgtatccctggcaccacaaatacaatcgaatcggagcatttcatacgtgtgaggact 389
Qy 390 cgtgctgttgcctttcatggatcgctatcggaagctcagtgaccatgatacctgctcaaca 449
Db 390 cgtgctgttgcctttcatggatcgctatcggaagctcagtgaccatgatacctgctcaaca 449
Qy 450 gatcccggtcgctgatagccgcttatagtacacatactcccaggattcgtggcatgtctt 509
Db 450 gatcccggtcgctgatagccgcttatagtacacatactcccaggattcgtggcatgtctt 509
Qy 510 cctcatctacgagggtcgctggtggtctctgtttcaactgtttgcccctgaaaaag 569
Db 510 cctcatctacgagggtcgctggtggtctctgtttcaactgtttgcccctgaaaaag 569
Qy 570 aaaccttgggttcataaatacggatccggctcagcatcgctctctcgtgatctcctt 629
Db 570 aaaccttgggttcataaatacggatccggctcagcatcgctctctcgtgatctcctt 629
Qy 630 tatcgccattctagcgggtccaaaccccaaggctccaaactcacaggtatggactgcttg 689
Db 630 tatcgccattctagcgggtccaaaccccaaggctccaaactcacaggtatggactgcttg 689
Qy 690 gagcaactatactggtggtccgacgggtctgcttcatctcctgggcttctcgacatcctg 749
Db 690 gagcaactatactggtggtccgacgggtctgcttcatctcctgggcttctcgacatcctg 749
Qy 750 ctctcatgttcatggtggtgacgcagcaatgcacatctggtggtgaagaatgcacagatgctgc 809
Db 750 ctctcatgttcatggtggtgacgcagcaatgcacatctggtggtgaagaatgcacagatgctgc 809
Qy 810 tegtacggttaccccaagcagtggttcagtgaataataatgtgcttctgcaccgcctttcc 869
Db 810 tegtacggttaccccaagcagtggttcagtgaataataatgtgcttctgcaccgcctttcc 869
Qy 870 atatacaatcgcagttctgtatggaaattacagatctcgcactctattcttaagtctccgg 929
Db 870 atatacaatcgcagttctgtatggaaattacagatctcgcactctattcttaagtctccgg 929
Qy 930 ctatatctccattccagacaaatgacgcagtcctctcgtgctcagtttttgcaacggtcct 989
Db 930 ctatatctccattccagacaaatgagggcagctctgctcgaatttggcctttttgcaacggtcct 989
Qy 990 ctcatgtggcgggtatcgtgatggccttcttgcctcctcaacgctgtacaagagactgcgtc 1049
Db 990 ctcatgtggcgggtatcgtgatggccttcttgcctcctcaacgctgtacaagagactgcgtc 1049
Qy 1050 tcgactcacctggagccttggccggggaataatgggctgtattttccactcatctcgaacg 1109
Db 1050 tcgactcacctggagccttggccggggaataatgggctgtattttccactcatctcgaacg 1109
Qy 1110 cattcatccccgcgtggcaagtctcgtgttgggtctctatttcggacacctggggaattctggc 1169
Db 1110 cattcatccccgcgtggcaagtctcgtgttgggtctctatttcggacacctggggaattctggc 1169
Qy 1170 cacatgcggatgtatatcttaggttagcacagctttcaatgccttggtcaattccgc 1229
Db 1170 cacatgcggatgtatatcttaggttagcacagctttcaatgccttggtcaattccgc 1229
Qy 1230 cgttgctaccagaactctccttctgatcccaatcgccctactcctctacacaaagcg 1289

Db 1230 cgttgctaccagaactctccttctgatcccaatcgccctactcctctacacaaagcg 1289
Qy 1290 agatccaaagtcttgcggagcactcgtgctttgtttaccgctgggaatcgggtttct 1349
Db 1290 agatccaaagtcttgcggagcactcgtgctttgtttaccgctgggaatcgggtttct 1349
Qy 1350 ggtcaatgtcgtagcgtgtgttcttcacgtccgtccaccactgtgtttttcagcttcccact 1409
Db 1350 ggtcaatgtcgtagcgtgtgttcttcacgtccgtccaccactgtgtttttcagcttcccact 1409
Qy 1410 gaccgtgcctacggccgcgtcaaccatgaattacacaaagtgcgattataggcgttgcaact 1469
Db 1410 gaccgtgcctacggccgcgtcaaccatgaattacacaaagtgcgattataggcgttgcaact 1469
Qy 1470 tgccttgggtgtcttgaactgggtcgtgcatgccaggaagcattatccagggacccactt 1529
Db 1470 tgccttgggtgtcttgaactgggtcgtgcatgccaggaagcattatccagggacccactt 1529
Qy 1530 ggagcttgacgggacgggtcgtcggagcagaatttcaagtgggccaatga 1578
Db 1530 ggagcttgacgggacgggtcgtcggagcagaatttcaagtgggccaatga 1578

RESULT 14
US-09-351-823-7
; Sequence 7, Application US/09351823
; GENERAL INFORMATION:
; APPLICANT: DuVick, Jon
; TITLE OF INVENTION: Compositions and Methods for Fummonisin Detoxification
; FILE REFERENCE: 5718-56
; CURRENT APPLICATION NUMBER: US/09/351,823
; CURRENT FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 1578
; TYPE: DNA
; ORGANISM: Exophiala spinifera
; FEATURE:
; OTHER INFORMATION: permease, coding sequence
US-09-351-823-7

Query Match 59.4%; Score 937; DB 17; Length 1578;
Best Local Similarity 99.2%; Pred. No. 0;
Matches 1537; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 30 gaaaggcgggacaaaggcagacaaacgaagaacacagagacggcggcgagcaggtggtgcgtc 89
Db 30 gaaaggcgggacaaaggcagacaaacgaagaacacagagacggcggcgagcaggtggtgcgtc 89
Qy 90 cgaagtcctgaacgttctctctggaagaagaacaatttggcaccatcacctcgtgtcctt 149
Db 90 cgaagtcctgaacgttctctctggaagaagaacaatttggcaccatcacctcgtgtcctt 149
Qy 150 ggccttctgatttgcaacagttgggctggtatctcaggcagtcctccagctcgcctact 209
Db 150 ggccttctgatttgcaacagttgggctggtatctcaggcagtcctccagctcgcctact 209
Qy 210 agcggggggcccgctcactctcctttacggcatcctaatacagtaactcgtctacatctg 269
Db 210 agcggggggcccgctcactctcctttacggcatcctaatacagtaactcgtctacatctg 269
Qy 270 catcgccttctcattagccgaactgaccagcgtctacccgactgccggtggccaatatca 329
Db 270 catcgccttctcattagccgaactgaccagcgtctacccgactgccggtggccaatatca 329
Qy 330 ttttgctcgtatccctggcaccacaaatacaatcaatcggagcatttcatacgtgtgaggact 389
Db 330 ttttgctcgtatccctggcaccacaaatacaatcaatcggagcatttcatacgtgtgaggact 389
Qy 390 cgttgctaccagaactctccttctgatcccaatcgccctactcctctacacaaagcg 449

Db 390 cgtgctgttcttcatggatcgctctatcggaagctcagtgaccatgatacctgctcaaca 449
Qy 450 gatcccgccgtgatagccgcttatagtcacacatactccaggattcgtgcatgtctt 509
Db 450 gatcccgccgtgatagccgcttatagtcacacatactccaggattcgtgcatgtctt 509
Qy 510 cctcatctacaggggagtcgctgctgtgctcttgttcaacttgtttgcccctgaaaaa 569
Db 510 cctcatctacaggggagtcgctgctgtgctcttgttcaacttgtttgcccctgaaaaa 569
Qy 570 aaacccttgggttcataaatacggattcggcctcagatcgctctctcttgatctcctt 629
Db 570 aaacccttgggttcataaatacggattcggcctcagatcgctctctcttgatctcctt 629
Qy 630 tatcgccattctagcgcggtcccaaccccccaaggctccaaactcaccaggtatgactgctt 689
Db 630 tatcgccattctagcgcggtcccaaccccccaaggctccaaactcaccaggtatgactgctt 689
Qy 690 gagcaactatactaggctggtccgacggcgtctgcttcatcctggcctttcgacatcctg 749
Db 690 gagcaactatactaggctggtccgacggcgtctgcttcatcctggcctttcgacatcctg 749
Qy 750 ctctcatgttcataggcttggacgcagcaaatgcatctctggtggaagaatgcacagatgctgc 809
Db 750 ctctcatgttcataggcttggacgcagcaaatgcatctctggtggaagaatgcacagatgctgc 809
Qy 810 tcgtacggtaccacaaagcagtggtcagtgcaatcataatggcttctctgcaccgcttctc 869
Db 810 tcgtacggtaccacaaagcagtggtcagtgcaatcataatggcttctctgcaccgcttctc 869
Qy 870 atatacaatcgagcttctgtatggattacagatctcgactctattctaaagtccgcgg 929
Db 870 atatacaatcgagcttctgtatggattacagatctcgactctattctaaagtccgcgg 929
Qy 930 ctatatccattcagacaaatgacgcaglcctctcgtcgtcagttttgcaacgggtcct 989
Db 930 ctatatccattcagacaaatgagcgagctcgtcgtcgaattcgtcttttgcacgggtcct 989
Qy 990 ctcatgtggcgttatcgtgatggcctctcttcgcccctcaacgctgtacaagagactgcgtc 1049
Db 990 ctcatgtggcgttatcgtgatggcctctcttcgcccctcaacgctgtacaagagactgcgtc 1049
Qy 1050 tcgactcacctggagcttggcccggaacaatggcgtggtattttccactcactcgaacg 1109
Db 1050 tcgactcacctggagcttggcccggaacaatggcgtggtattttccactcactcgaacg 1109
Qy 1110 cattcatcccgctggcaagtctcctgtttggtctctatttcgcacctggggaattctgyc 1169
Db 1110 cattcatcccgctggcaagtctcctgtttggtctctatttcgcacctggggaattctgyc 1169
Qy 1170 cacatgcggatgtatatcttaggttttagcacacagctttcaatgccttgggtcaattccgc 1229
Db 1170 cacatgcggatgtatatcttaggttttagcacacagctttcaatgccttgggtcaattccgc 1229
Qy 1230 cgttgtactccagcaactctccttctcgtgatcccaatcgccctactcctctacacaaagcg 1289
Db 1230 cgttgtactccagcaactctccttctcgtgatcccaatcgccctactcctctacacaaagcg 1289
Qy 1290 agatccaaagtcttgcgagcactcgtgtttgtgttaccgctgggaatcgggtttct 1349
Db 1290 agatccaaagtcttgcgagcactcgtgtttgtgttaccgctgggaatcgggtttct 1349
Qy 1350 ggtcaaatgtcgtagcgggtgttcttcaacgtccgctcaccactgtgttttccagcttcccaact 1409
Db 1350 ggtcaaatgtcgtagcgggtgttcttcaacgtccgctcaccactgtgttttccagcttcccaact 1409
Qy 1410 gaccgtgccttacgcccgcgtcaaccatgaattacacaaagtgcgattataggcgttgcact 1469
Db 1410 gaccgtgccttacgcccgcgtcaaccatgaattacacaaagtgcgattataggcgttgcact 1469
Qy 1470 tgccttgggtgtcttgaactgggtcgtgcatgccagggaagcaattatcagggaccacactt 1529
Db 1470 tgccttgggtgtcttgaactgggtcgtgcatgccagggaagcaattatcagggaccacactt 1529

Qy 1530 ggagcttgacggacgggtcgtcgagcagaatttcaagtgggcatga 1578
Db 1530 ggagcttgacggacgggtcgtcgagcagaatttcaagtgggcatga 1578
RESULT 15
US-09-677-488-7
; Sequence 7, Application US/09677488
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jon
; TITLE OF INVENTION: Compositions and Methods for Fummonisin Detoxification
; FILE REFERENCE: 35718/204100 (5718-111A)
; CURRENT APPLICATION NUMBER: US/09/677,488
; CURRENT FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 1578
; TYPE: DNA
; ORGANISM: Exophiala spinifera
; FEATURE:
; OTHER INFORMATION: permease, fully spliced cDNA
US-09-677-488-7

Query Match 59.4%; Score 937; DB 26; Length 1578;
Best Local Similarity 99.2%; Pred. No. 0;
Matches 1537; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
Qy 30 gaaagcgcggaacagggcagacaaacgaagaacacagagacggcgggcaggtggtgcgc 89
Db 30 gaaagcgcggaacagggcagacaaacgaagaacacagagacggcgggcaggtggtgcgc 89
Qy 90 cgaagtcctgaacgttctctgtggaagaacaaatttggcaccatcaccatcgtgctct 149
Db 90 cgaagtcctgaacgttctctgtggaagaacaaatttggcaccatcaccatcgtgctct 149
Qy 150 ggcccttctgatttgcacaaagttgggctggtatctcagggcagctcctcagctcgcctact 209
Db 150 ggcccttctgatttgcacaaagttgggctggtatctcagggcagctcctcagctcgcctact 209
Qy 210 agcggggggggcccgctcactctcctttacggcgaactcactaatcagctactcgtctacatctg 269
Db 210 agcggggggggcccgctcactctcctttacggcgaactcactaatcagctactcgtctacatctg 269
Qy 270 catcgcttctcattagccgaactgaccagcgtctacccgactgcccgggtggcccaatata 329
Db 270 catcgcttctcattagccgaactgaccagcgtctacccgactgcccgggtggcccaatata 329
Qy 330 ttttgcgtcgtatcctcctggcaccacaaatacaatcaatcgaatcagcatttccatacgtgtgcggact 389
Db 330 ttttgcgtcgtatcctcctggcaccacaaatacaatcaatcgaatcagcatttccatacgtgtgcggact 389
Qy 390 cgtgtcgttgccttctcatggatcgctatcggaagctcagtgaccatgatacctgctcaaca 449
Db 390 cgtgtcgttgccttctcatggatcgctatcggaagctcagtgaccatgatacctgctcaaca 449
Qy 450 gatccggcgctgtagcgcgcctatagtcacacatactcccaggattcgttggcctgctt 509
Db 450 gatccggcgctgtagcgcgcctatagtcacacatactcccaggattcgttggcctgctt 509
Qy 510 cctcatctacaggggagtcgctggtggtgctcttcttgcacttgcctgctgctgctgctt 569
Db 510 cctcatctacaggggagtcgctggtggtgctcttcttgcacttgcctgctgctgctgctt 569
Qy 570 aaacccttgggttcataaatacggattcggcctcagcagctcgtctctctcgtgactcctt 629
Db 570 aaacccttgggttcataaatacggattcggcctcagcagctcgtctctctcgtgactcctt 629
Qy 630 tatcgccattctagcgcggtcccaaccccccaaggctccaaactcaccaggtatgactgctt 689

||||| 630 tatcgccattctatgcggttccaaaccccaaggctccaaactcacaggtatggactgcttg 689

Qy 690 gagcaactatactggtggtccgacggcgctgctgttcactcctggccttttcacacatcctg 749

Db 690 gagcaactatactggtggtccgacggcgctgctgttcactcctggccttttcacacatcctg 749

Qy 750 cttcatgttcatctggttggacgcgaacaatgcattcgtggtgaagaatgcacagatgctgc 809

Db 750 cttcatgttcatctggttggacgcgaacaatgcattcgtggtgaagaatgcacagatgctgc 809

Qy 810 tegtacgggtacccaaagcagtggtcagtgcaataataatggcttctgcacgcgcctttcc 869

Db 810 tegtacgggtacccaaagcagtggtcagtgcaataataatggcttctgcacgcgcctttcc 869

Qy 870 atatacaaatcgcagttctgttatggaaattacagatctcgaactctattctaaagtccgcgg 929

Db 870 atatacaaatcgcagttctgttatggaaattacagatctcgaactctattctaaagtccgcgg 929

Qy 930 ctatatattccattccgagacaaatgacgcagtcaccttcggtcggctcagtttttgcaacggctcct 989

Db 930 ctatatattccattccgagacaaatgaggaagtcctgctcgaaattcggctttttgcaacggctcct 989

Qy 990 ctoaatgtggcggtatcgtgatggccttcttcgcacctcaacgctgtacaagagagactgcgtc 1049

Db 990 ctoaatgtggcggtatcgtgatggccttcttcgcacctcaacgctgtacaagagagactgcgtc 1049

Qy 1050 tcgactcacctggagctttgcccgggacaatggcgtggtattttccactcatctcgaacg 1109

Db 1050 tcgactcacctggagctttgcccgggacaatggcgtggtattttccactcatctcgaacg 1109

Qy 1110 cattcatcccgctggcaagttcctgttggctctctattcgcgcacctggggaattcttggc 1169

Db 1110 cattcatcccgctggcaagttcctgttggctctctattcgcgcacctggggaattcttggc 1169

Qy 1170 cacatgcggtatgtatatattcttagttctagcacagctttcaatgccttggtcaattccgc 1229

Db 1170 cacatgcggtatgtatatattcttagttctagcacagctttcaatgccttggtcaattccgc 1229

Qy 1230 cgttgtaactccagcaactctcctctccgtatcccaatcgcctactcctctacccaaaagcg 1289

Db 1230 cgttgtaactccagcaactctcctctccgtatcccaatcgcctactcctctacccaaaagcg 1289

Qy 1290 agatccaaaagtcttgcgcgagcactcgtcttttggttaccgcggtggaatcggggtttct 1349

Db 1290 agatccaaaagtcttgcgcgagcactcgtcttttggttaccgcggtggaatcggggtttct 1349

Qy 1350 ggtcaaatgtgctagcgggtggtcttccagtcgcgtcaccactggtttttcagcttcccaact 1409

Db 1350 ggtcaaatgtgctagcgggtggtcttccagtcgcgtcaccactggtttttcagcttcccaact 1409

Qy 1410 gaccgtgcctacggccgcggtcaaccatgaattacacaagtgcgattatagcggttgcaact 1469

Db 1410 gaccgtgcctacggccgcggtcaaccatgaattacacaagtgcgattatagcggttgcaact 1469

Qy 1470 tgcctcttgggtgcttgaactgggtcgtgcatgccagggaagcattatcaggggaccccaact 1529

Db 1470 tgcctcttgggtgcttgaactgggtcgtgcatgccagggaagcattatcaggggaccccaact 1529

Qy 1530 ggagcttgacggacggggtcgtcggagcagaatttcaagttcgggccatga 1578

Db 1530 ggagcttgacggacggggtcgtcggagcagaatttcaagttcgggccatga 1578

RESULT 2
US-08-781-891-79/c
; Sequence 79, Application US/08781891
; Patent No. 6090620
; GENERAL INFORMATION:
; APPLICANT: Fu, Ying-Hui
; APPLICANT: Yu, Chang-En
; APPLICANT: Oshima, Junko
; APPLICANT: Mulligan, John T.
; APPLICANT: Schellenberg, Gerald D.
; TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
; TITLE OF INVENTION: WERNER'S SYNDROME
; NUMBER OF SEQUENCES: 209
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,891
; FILING DATE: 27-DEC-1996
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6090620tenburg Ph.D., Carol
; REGISTRATION NUMBER: 39,317
; REFERENCE/DOCKET NUMBER: 240052.419
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 79:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 87350 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-781-891-79

Query Match 1.2%; Score 19; DB 3; Length 87350;
Best Local similarity 100.0%; Pred. No. 3.6;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1471 gctcttggtgtcttgaact 1489
|||||
Db 57973 GCTCTTGGTGTCTTGAAC 57955

RESULT 3
US-07-745-382-21/c
; Sequence 21, Application US/07745382
; Patent No. 5270181
; GENERAL INFORMATION:
; APPLICANT: McCoy, John
; APPLICANT: Lavallie, Edward
; TITLE OF INVENTION: Peptide and Protein Fusions To
; TITLE OF INVENTION: Thioredoxin and Thioredoxin-Like Molecules
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/745,382
; FILING DATE: 19910814
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/652,531
; FILING DATE: 06-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseir, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: G15188A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 876-1170
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 327 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..327
; OTHER INFORMATION: /citation= (11)
; PUBLICATION INFORMATION:
; AUTHORS: Lim,
; JOURNAL: J. Bacteriol.
; VOLUME: 163
; PAGES: 311-316
; DATE: 1985
; RELEVANT RESIDUES IN SEQ ID NO: 21: FROM 1 TO 327
; US-07-745-382-21

Query Match 1.1%; Score 17; DB 1; Length 327;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 971 tcagttttgcaacggtc 987
|||||
Db 178 TCAGTTTGTCAACGGTC 162

RESULT 4
US-07-921-848-21/c
; Sequence 21, Application US/07921848
; Patent No. 5292646
; GENERAL INFORMATION:
; APPLICANT: McCoy, John
; APPLICANT: Lavallie, Edward
; TITLE OF INVENTION: Peptide and Protein Fusions To
; TITLE OF INVENTION: Thioredoxin and Thioredoxin-Like Molecules
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/921,848
; FILING DATE: 19920728

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/652,531
FILING DATE: 06-FEB-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/745,382
FILING DATE: 14-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: Cserr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI5188A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 327 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..327
OTHER INFORMATION: /citation= ([1])
PUBLICATION INFORMATION:
AUTHORS: Lim,
JOURNAL: J. Bacteriol.
VOLUME: 163
PAGES: 311-316
DATE: 1985
RELEVANT RESIDUES IN SEQ ID NO: 21: FROM 1 TO 327
US-07-921-848-21

Query Match 1.1%; Score 17; DB 1; Length 327;
Best Local Similarity 100.0%; pred. No. 35;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 971 tcagttttgcaacggtc 987
|||||

Db 178 TCAGTTTTCACCGGTC 162

RESULT 5
US-08-165-301A-21/c
Sequence 21, Application US/08165301A
Patent No. 5646016
GENERAL INFORMATION:
APPLICANT: McCoy, John
APPLICANT: DiBlasio-Smith, Elizabeth
APPLICANT: Grant, Kathleen
APPLICANT: Lavallie, Edward R.
TITLE OF INVENTION: PEPTIDE AND PROTEIN FUSIONS TO
THIOREDOXIN, THIOREDOXIN-LIKE MOLECULES, AND MODIFIED
THIOREDOXIN-LIKE MOLECULES
TITLE OF INVENTION: THIOREDOXIN-LIKE MOLECULES
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/165,301A
FILING DATE:
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Meinert, M. C.
REGISTRATION NUMBER: 33,544
REFERENCE/DOCKET NUMBER: GI 5188D
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 327 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..327
US-08-165-301A-21

Query Match 1.1%; Score 17; DB 1; Length 327;
Best Local Similarity 100.0%; pred. No. 35;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 971 tcagttttgcaacggtc 987
|||||

Db 178 TCAGTTTTCACCGGTC 162

RESULT 6
US-08-810-436-21/c
Sequence 21, Application US/08810436
Patent No. 6143524
GENERAL INFORMATION:
APPLICANT: McCoy, John
APPLICANT: DiBlasio-Smith, Elizabeth
APPLICANT: Grant, Kathleen
APPLICANT: Lavallie, Edward R.
TITLE OF INVENTION: PEPTIDE AND PROTEIN FUSIONS TO
THIOREDOXIN, THIOREDOXIN-LIKE MOLECULES, AND MODIFIED
THIOREDOXIN-LIKE MOLECULES
TITLE OF INVENTION: THIOREDOXIN-LIKE MOLECULES
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/810,436
FILING DATE: 04-MAR-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/165,301
FILING DATE: 10-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: Meinert, M. C.
REGISTRATION NUMBER: 33,544
REFERENCE/DOCKET NUMBER: GI 5188D
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 327 base pairs
TYPE: nucleic acid
STRANDEDNESS: double

ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/165,301A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Meinert, M. C.
REGISTRATION NUMBER: 33,544
REFERENCE/DOCKET NUMBER: GI 5188D
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 330 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..330
US-08-165-301A-27

Query Match 1.1%; Score 17; DB 1; Length 330;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 971 tcagttttgcaacggtc 987
|||||

Db 178 TCAGTTTTCGAACGGTC 162

RESULT 10

US-08-810-436-25/c
Sequence 25, Application US/08810436
Patent No. 6143524
GENERAL INFORMATION:
APPLICANT: McCoy, John
APPLICANT: DiBlasio-Smith, Elizabeth
APPLICANT: Grant, Kathleen
APPLICANT: Lavallie, Edward R.
TITLE OF INVENTION: PEPTIDE AND PROTEIN FUSIONS TO
THIOREDOXIN, THIOREDOXIN-LIKE MOLECULES, AND MODIFIED
THIOREDOXIN-LIKE MOLECULES
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/810,436
FILING DATE: 04-MAR-1997
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/165,301
FILING DATE: 10-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: Meinert, M. C.
REGISTRATION NUMBER: 33,544
REFERENCE/DOCKET NUMBER: GI 5188D
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 330 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..330
US-08-810-436-25

Query Match 1.1%; Score 17; DB 3; Length 330;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 971 tcagttttgcaacggtc 987
|||||

Db 178 TCAGTTTTCGAACGGTC 162

RESULT 11

US-08-810-436-27/c
Sequence 27, Application US/08810436
Patent No. 6143524
GENERAL INFORMATION:
APPLICANT: McCoy, John
APPLICANT: DiBlasio-Smith, Elizabeth
APPLICANT: Grant, Kathleen
APPLICANT: Lavallie, Edward R.
TITLE OF INVENTION: PEPTIDE AND PROTEIN FUSIONS TO
THIOREDOXIN, THIOREDOXIN-LIKE MOLECULES, AND MODIFIED
THIOREDOXIN-LIKE MOLECULES
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/810,436
FILING DATE: 04-MAR-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/165,301
FILING DATE: 10-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: Meinert, M. C.
REGISTRATION NUMBER: 33,544
REFERENCE/DOCKET NUMBER: GI 5188D
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:

; LENGTH: 330 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 1..330
 ;
 US-08-810-436-27

Query Match 1.1%; Score 17; DB 3; Length 330;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 971 tcagttttgcaacgggc 987
 |||||
 Db 178 TCAGTTTTCACACGGTC 162

RESULT 12
 PCT-US94-14179-25/c
 ; Sequence 25, Application PC/TUS9414179
 ; GENERAL INFORMATION:
 ; APPLICANT: McCoy, John
 ; APPLICANT: DiBlasio-Smith, Elizabeth
 ; APPLICANT: Grant, Kathleen
 ; APPLICANT: Lavallie, Edward R.
 ; TITLE OF INVENTION: PEPTIDE AND PROTEIN FUSIONS TO
 ; TITLE OF INVENTION: THIOREDOXIN, THIOREDOXIN-LIKE MOLECULES, AND MODIFIED
 ; TITLE OF INVENTION: THIOREDOXIN-LIKE MOLECULES
 ; NUMBER OF SEQUENCES: 29
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Genetics Institute, Inc.
 ; STREET: 87 CambridgePark Drive
 ; CITY: Cambridge
 ; STATE: Massachusetts
 ; COUNTRY: U.S.A.
 ; ZIP: 02140
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US94/14179
 ; FILING DATE:
 ; CLASSIFICATION:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Meinert, M. C.
 ; REGISTRATION NUMBER: 33,544
 ; REFERENCE/DOCKET NUMBER: GI 5188D
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617) 876-1170
 ; TELEFAX: (617) 876-5851
 ; INFORMATION FOR SEQ ID NO: 25:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 330 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 1..330
 ;
 PCT-US94-14179-25

Query Match 1.1%; Score 17; DB 5; Length 330;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 971 tcagttttgcaacgggc 987

Db 178 TCAGTTTTCACACGGTC 162
 |||||

RESULT 13
 PCT-US94-14179-27/c
 ; Sequence 27, Application PC/TUS9414179
 ; GENERAL INFORMATION:
 ; APPLICANT: McCoy, John
 ; APPLICANT: DiBlasio-Smith, Elizabeth
 ; APPLICANT: Grant, Kathleen
 ; APPLICANT: Lavallie, Edward R.
 ; TITLE OF INVENTION: PEPTIDE AND PROTEIN FUSIONS TO
 ; TITLE OF INVENTION: THIOREDOXIN, THIOREDOXIN-LIKE MOLECULES, AND MODIFIED
 ; TITLE OF INVENTION: THIOREDOXIN-LIKE MOLECULES
 ; NUMBER OF SEQUENCES: 29
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Genetics Institute, Inc.
 ; STREET: 87 CambridgePark Drive
 ; CITY: Cambridge
 ; STATE: Massachusetts
 ; COUNTRY: U.S.A.
 ; ZIP: 02140
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US94/14179
 ; FILING DATE:
 ; CLASSIFICATION:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Meinert, M. C.
 ; REGISTRATION NUMBER: 33,544
 ; REFERENCE/DOCKET NUMBER: GI 5188D
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617) 876-1170
 ; TELEFAX: (617) 876-5851
 ; INFORMATION FOR SEQ ID NO: 27:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 330 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 1..330
 ;
 PCT-US94-14179-27

Query Match 1.1%; Score 17; DB 5; Length 330;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 971 tcagttttgcaacgggc 987
 |||||
 Db 178 TCAGTTTTCACACGGTC 162

RESULT 14
 US-08-683-007A-1/c
 ; Sequence 1, Application US/08683007A
 ; Patent No. 5858724
 ; GENERAL INFORMATION:
 ; APPLICANT: NO. 5858724y, Robert E
 ; APPLICANT: Domanico, Michael
 ; APPLICANT: Yeager, Keith
 ; APPLICANT: Kroeker, Warren
 ; TITLE OF INVENTION: Recombinant Rabbit Tissue Factor
 ; NUMBER OF SEQUENCES: 5
 ; CORRESPONDENCE ADDRESS:

ADDRESSEE: Quarles & Brady
STREET: 1 South Pinckney Street
CITY: Madison
STATE: WI
COUNTRY: USA
ZIP: 53703
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/683,007A
FILING DATE: 16-JUL-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J
REGISTRATION NUMBER: 27386
REFERENCE/DOCKET NUMBER: 740380.90040
TELECOMMUNICATION INFORMATION:
TELEPHONE: 608-251-5000
TELEFAX: 608-251-9166
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1227 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "Recombinant gene encoding
DESCRIPTION: trxA-rabbit tissue factor fusion protein"
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1227
US-08-683-007A-1

Query Match 1.1%; Score 17; DB 2; Length 1227;
Best Local Similarity 100.0%; Pred.No. 36;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 971 tcagttttgcaacggtc 987
|||||
Db 178 TCAGTTTTGCAACGGTC 162

RESULT 15
US-09-237-111-1/c
Sequence 1, Application US/09237111
Patent No. 6340564
GENERAL INFORMATION:
APPLICANT: Zalacain, Magdalena
APPLICANT: Burnham, Martin K. R.
APPLICANT: Biswas, Sanjoy
APPLICANT: Brown, James R.
APPLICANT: Warren, Patrick V.
APPLICANT: Ingraham, Karen A.
APPLICANT: Chalker, Alison F.
APPLICANT: So, Chi-Young
APPLICANT: Holmes, David J.
APPLICANT: Van Horn, Stephanie
APPLICANT: Warren, Richard L.
TITLE OF INVENTION: YhxB
FILE REFERENCE: GM10190
CURRENT APPLICATION NUMBER: US/09/237,111
CURRENT FILING DATE: 1999-01-26
NUMBER OF SEQ ID NOS: 2
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1
LENGTH: 1719
TYPE: DNA
ORGANISM: Streptococcus pneumoniae
US-09-237-111-1

Query Match 1.1%; Score 17; DB 4; Length 1719;
Best Local Similarity 100.0%; Pred.No. 36;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 427 gtgaccatgatacctgc 443
|||||
Db 425 GTGACCATGATACCTGC 409

Search completed: April 27, 2002, 05:55:15
Job time: 15402 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2002, 05:28:23 ; Search time 705 seconds
(without alignments)
3842.966 Million cell updates/sec

Title: US-09-882-694A-7
Perfect score: 1578
Sequence: 1 atggactccagaccagtg.....aatttcaagttggccatga 1578

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1736436 seqs, 858457221 residues

Word size : 0
Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	937	59.4	1764	AAD02693	Exophiala spinifera
C 2	21	1.3	11866	AA040141	Sequence of human
C 3	21	1.3	11873	AA040176	Sequence of human
C 4	21	1.3	38059	AAF54018	Human factor IX (h
5	20	1.3	390	AAC02802	Human secreted pro
6	19	1.2	379	AAT12188	pJG4-5-CDK-BP cDNA
7	19	1.2	718	ABL18293	Drosophila melanog
8	19	1.2	1170	AAF63760	Drosophila gustato
C 9	19	1.2	1254	AAT31455	CDK6 inhibiting pr

C 10	19	1.2	2302	22	AAK83590	Human immune/haema
C 11	19	1.2	2718	23	ABL18292	Drosophila melanog
C 12	19	1.2	4519	23	ABL22926	Drosophila melanog
13	19	1.2	12332	23	ABL10880	Drosophila melanog
14	19	1.2	27869	22	ABA19635	Human nervous syst
C 15	19	1.2	27869	22	AAK66517	Human immune/haema
C 16	19	1.2	47670	23	ABL16824	Drosophila melanog
C 17	19	1.2	50368	23	ABL16768	Drosophila melanog
C 18	19	1.2	87350	18	AAK83003	Human WRN genomic
C 19	18	1.1	467	21	AAA89575	Exol9 nucleotide s
20	18	1.1	555	19	AAV16456	Endo-beta-1,4-gluc
C 21	18	1.1	602	22	AAH87728	Peppermint plant o
C 22	18	1.1	897	24	ABI99515	Mouse ischaemic co
C 23	18	1.1	969	24	AAS97221	Neisseria meningit
24	18	1.1	1219	21	AAA89551	Mouse syntaxin4 nu
C 25	18	1.1	1553	23	AAS86570	DNA encoding novel
C 26	18	1.1	1596	17	AAK22316	Nocardiosis sp. p
27	18	1.1	3000	21	AAK65523	Porcine BAC-PiGf2-
28	18	1.1	3078	20	AAK13360	Enterococcus faeca
29	18	1.1	3401	23	ABL02901	Drosophila melanog
30	18	1.1	3447	19	AAV16444	Glucanase II gene
31	18	1.1	5421	23	AAS75392	DNA encoding novel
C 32	18	1.1	6007	23	ABL02900	Drosophila melanog
C 33	18	1.1	6080	23	AAS72880	DNA encoding novel
C 34	18	1.1	168575	22	AAH21613	Human hypodermis r
C 35	17	1.1	193	22	AAK70598	Human immune/haema
36	17	1.1	210	22	AAH66973	C glutamic acid codin
37	17	1.1	229	20	AAV87341	EST clone BW51. H
38	17	1.1	249	19	AAV03550	DNA sequence that
C 39	17	1.1	276	19	AAV32676	Delta thioresoxin
C 40	17	1.1	280	19	AAV32671	TrxA-concat sequen
C 41	17	1.1	282	19	AAV32672	Delta thioresoxin
C 42	17	1.1	323	22	AAH12667	Human cDNA clone (
C 43	17	1.1	327	16	AAQ90778	E. coli thioresoxi
C 44	17	1.1	329	16	AAQ90775	E. coli hpTRX gene
C 45	17	1.1	329	16	AAQ90776	E. coli hp2TRX gen

ALIGNMENTS

RESULT 1

AAD02693
ID AAD02693 standard; DNA; 1764 BP.

XX AAD02693;

XX 02-MAY-2001 (first entry)

XX Exophiala spinifera permease DNA.

XX Fumonisin; degradative enzyme; transport enzyme; flavin monooxygenase;
KW aldehyde dehydrogenase; permease; p-glycoprotein; fumonisin esterase;
KW detoxification; mycotoxin; animal feed; human feed; silage;
KW transgenic plant; transgenic animal; microbial spray; ds.

OS Exophiala spinifera.

XX Key Location/Qualifiers

FT CDS 5..1582

FT /tag= a

FT /product= "E. spinifera permease"

FT /note= "This region is specifically claimed in

FT claim 1b as SEQ ID NO: 7"

XX WO200105980-A1.

XX 25-JAN-2001.

XX 14-JUL-1999; 99WO-US15824.

XX 14-JUL-1999; 99WO-US15824.

(PION-) PIONEER HI-BRED INT INC.
(CURA-) CURAGEN CORP.
Duvick JP, Maddox J, Gilliam J, Folkerts O, Crasta OR;
WPI; 2001-147345/15.
P-PSDB; AAY72636.
Novel polynucleotides encoding Exophiala degradative or transport enzyme which is useful for detoxifying fumonisin or structurally related mycotoxin during processing of grain for human or animal food consumption -
Claim 1b; Page 70; 90pp; English.
The patent discloses novel polynucleotides encoding Exophiala spinifera fumonisin degradative or transport enzymes such as flavin monooxygenase, aldehyde dehydrogenase, permease and p-glycoprotein. A fumonisin esterase enzyme with at least one fumonisin degradative enzyme is useful for detoxifying fumonisin or a structurally related mycotoxin present in harvested grain, on application to a plant or to harvested grain during processing, or to processed grain that is to be used as animal or human feed, or as a silage. Nucleotide fragments of the present invention are useful as probes and primers. They can be introduced into microorganisms that multiply on plants to deliver enzymes to potential target crops. The genes encoding the degrading enzymes are introduced via a vector into a microbial host and the transformed host is supplied to the environment, plants or animals for reducing the pathogenicity of a fungus producing fumonisin. The genes of the invention are fermented in a bacterial host and the resulting bacteria is processed and used as a microbial spray. The nucleotide sequences can be used alone or in combination to engineer microbes or other organisms to metabolise fumonisin and resist its toxic effects. The present sequence is Exophiala spinifera DNA encoding permease, a fumonisin degradative enzyme.
Sequence 1764 BP; 397 A; 478 C; 432 G; 454 T; 3 other;

	Query Match	59.4%;	Score 937;	DB 22;	Length 1764;
	Best Local Similarity	99.2%;	Pred. NO. 0;		
	Matches 1537;	Conservative	0;	Mismatches 12;	Indels 0; Gaps 0;
QY	30	gaaaggcgagacaaggcgagacaacgaagaacacagagacggcgcgaggtggtgcgtc	89		
DB	34	gaaaggcgagacaaggcgagacaacgaagaacacagagacggcgcgaggtggtgcgtc	93		
QY	90	cgagtccctgaacgttccctctgagagaagaacaaatttggcaccatcaccatcggtgcctt	149		
DB	94	cgagtccctgaacgttccctctgagagaagaacaaatttggcaccatcaccatcggtgcctt	153		
QY	150	ggccttttgatttgcacaacagttgggctggtatctcaggcagtcctccagctcgccctact	209		
DB	154	ggccttttgatttgcacaacagttgggctggtatctcaggcagtcctccagctcgccctact	213		
QY	210	agcggggggggcccgctcactctctctttacgggcatacctcaatcagtcactctcgtctacatctg	269		
DB	214	agcggggggggcccgctcactctctctttacgggcatacctcaatcagtcactctcgtctacatctg	273		
QY	270	catcgctttctcattagccgaactgaccagcgctctaccggactgccgggtggccaataca	329		
DB	274	catcgctttctcattagccgaactgaccagcgctctaccggactgccgggtggccaataca	333		
QY	330	ttttgcgtcgatcctggcaccacaaatcaatcaa tggagcatttcatcacgtgtgcggact	389		
DB	334	ttttgcgtcgatcctggcaccacaaatcaatcaa tggagcatttcatcacgtgtgcggact	393		
QY	390	cgtgtcgttgcttttcattggtatcgctatcggaagctcagtgaccatgatacctgctcaaca	449		
DB	394	cgtgtcgttgcttttcattggtatcgctatcggaagctcagtgaccatgatacctgctcaaca	453		
QY	450	gatcccgcgctgatagccgcctatagtcacacatactcccaggattcgtggcattgtctt	509		

```
RESULT 2
AAN40141/c
ID AAN40141 standard; DNA; 11866 BP.
XX
AC AAN40141;
XX
DT 11-FEB-1992 (first entry)
XX
DE Sequence of human factor IX genomic DNA.
XX
KW Christmas disease; therapy; haemophilia; factor IX; blood clotting;
  diagnosis; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
  exon 4441..4560
  FT /*tag= a
  FT /label= AAP40692
  FT 7139..7342
  FT /*tag= b
  FT /label= AAP40693
XX
PN W08400560-A.
XX
PD 16-FEB-1984.
XX
PF 03-AUG-1983; 83WO-GB00191.
XX
PR 06-MAY-1983; 83GB-0012491.
PR 04-AUG-1982; 82GB-0022485.
XX
PA (NATR ) NATIONAL RES DEV CORP.
PA (BROW/) BROWNLEE G G.
XX
PI Brownlee G, Choo KH;
XX
WPI; 1984-049331/08.
DR P-PSDB; AAP40692,P40693.
XX
PT Recombinant DNA cloning vehicles - useful in prodn. of factor IX
  polypeptide and of diagnostic probes for Christmas disease
XX
PS Example; Fig 7a-i; 7lpp; English.
XX
CC The inventors claim DNA molecules comprising part or all of the
  human factor IX DNA. The invention also includes cDNA derived from
  human factor IX RNA. Specifically claimed are: recombinant DNA (the
  phage present in clone lambda HIX-1) deposited as NCIB No. 11749;
  CC Recombinant DNA in which the cloning vehicle is the modified pAT 153
  CC plasmid present in E.coli NCIB No. 11747; Recombinant DNA in which
  CC the bovine factor IX DNA sequence is contained in the recombinant
  CC DNA transformed into E.coli to form a clone deposited as NCIB No.
  CC 11748.
XX
SQ Sequence 11866 BP; 3616 A; 2082 C; 2165 G; 3180 T; 823 other;

Query Match 1.38; Score 21; DB 5; Length 11866;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 827 cagtggtcagtgcaatcataa 847
  |||||
Db 1286 CAGTGGTCAGTGCATCATATA 1266

RESULT 3
AAN40176/c
ID AAN40176 standard; DNA; 11873 BP.
XX
```

```
AC AAN40176;
XX
DT 13-FEB-1992 (first entry)
XX
DE Sequence of human factor IX genomic DNA.
XX
KW Haemophilia; Christmas disease; diagnosis; treatment; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
  CDS 4441..4560
  FT /*tag= a
  FT /note= "see AAP40220"
  FT 7139..7342
  FT /*tag= b
  FT /note= "see AAP40221"
  FT 4442..4570
  FT /*tag= c
  FT 7140..7342
  FT /*tag= d
  FT 7960..8155
  FT /*tag= e
  FT /rpt_type= AluI
  FT 9671..9938
  FT /*tag= f
  FT /rpt_type= AluI
XX
PN GB2125409-A.
XX
PD 07-MAR-1984.
XX
PF 03-AUG-1983; 83GB-0020975.
XX
PR 16-MAY-1983; 83GB-0012490.
PR 04-AUG-1982; 82GB-0022486.
PR 03-AUG-1983; 83GB-0020975.
XX
PA (NATR ) NATIONAL RES DEV CORP.
XX
PI Brownlee GG, Choo KH;
XX
WPI; 1984-057898/10.
DR P-PSDB; AAP40220,P40221.
XX
PT Prodn. of artificial human factor IX - by use of recombinant DNA
  sequences for host transformation and cultivation
XX
PS Example; Fig 7; 49pp; English.
XX
CC The inventors claim a recombinant DNA having a human factor IX
  sequence pref. at least 50 nucleotides long, esp. 75-27000
  CC nucleotides. A cloning vector contg. foreign DNA is also claimed.
  CC The foreign sequence pref. includes the whole of an exon sequence of
  CC the human factor IX genome. The cloning vehicle may be a modified
  CC pAT 153 plasmid. Also claimed is a labelled diagnostic probe
  CC comprising a DNA molecule having a single- or double-stranded probe
  CC sequence of 15 to 10000 nucleotides long Factor IX DNA sequence.
XX
SQ Sequence 11873 BP; 3615 A; 2078 C; 2170 G; 3180 T; 830 other;

Query Match 1.38; Score 21; DB 5; Length 11873;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 827 cagtggtcagtgcaatcataa 847
  |||||
Db 1286 CAGTGGTCAGTGCATCATATA 1266

RESULT 4
AAN4018/c
```


ID AAF54018 standard; DNA; 38059 BP.
XX AAF54018;
AC
XX 30-MAR-2001 (first entry)
DT
XX Human factor IX (hFIX) gene, SEQ ID NO:4.
DE
XX Age-related gene regulation; liver-specific; gene expression;
KW human factor IX; hFIX; AE5'; AE3'; age-regulatable expression construct;
KW antisense therapy; gene therapy; thrombosis; cardiovascular disease;
KW diabetes; Alzheimer's disease; Parkinson's disease; cancer; osteoporosis;
KW osteoarthritis; dementia; ds.
XX
OS Homo sapiens.
XX WO200075279-A2.
PN
XX 14-DEC-2000.
PD
XX 06-JUN-2000; 2000WO-US15728.
PF
XX 09-JUN-1999; 99US-0328925.
PR
XX (UNMI) UNIV MICHIGAN.
PA
XX Kurachi K, Kurachi S;
PI WPI; 2001-061708/07.
XX P-PSDB; AAB60281, AAB60282, AAB60283, AAB60284, AAB60285, AAB60286,
DR AAB60287, AAB60288, AAB60289.
DR
XX New regulatory elements that control age-related gene expression,
PT useful in gene therapy and for reducing Factor IX expression -
PT
XX Disclosure; Fig 8A-E; 225pp; English.
PS
XX The invention relates to nucleic acid sequences which regulate gene
CC expression in an age-related manner and/or in a liver-specific manner.
CC The invention identifies regions of the human factor IX (hFIX) gene, and
CC a region of the human protein C (hPC) gene, which are age-related
CC regulatory sequences. The hFIX age-related regulatory sequences are
CC designated AE5' (AAF54016) and AE3' (AAF54017) and are found in the 5'
CC UTR (at position 2164-2165 of AAF54018) and 3' UTR (at position
CC 34383-35655 of AAF54018) respectively. These elements act synergistically
CC to increase hFIX levels over the lifespan of an individual; however, they
CC can independently exert effects on hFIX mRNA in an age-related manner,
CC with AE5' acting to stabilise hFIX mRNA, and AE3' acting to increase hFIX
CC mRNA levels, over time. AE5' also directs liver-specific expression. The
CC hPC gene age-related regulatory sequence is found in the 5' UTR
CC (AAF54081), and contains two PEA-3 (polyoma virus activator 3) elements
CC 5'-GAGGAAA-3' and 5'-CAGGAG-3'. The age-related regulatory sequences of
CC the invention, along with their homologues, variants and fragments, may
CC be used in the construction of recombinant expression vectors for the
CC expression of a desired sequence in an age-related fashion in a host
CC cell. Preferred target genes for expression in such age-regulatable
CC expression vectors include those encoding proteins involved in blood
CC coagulation (e.g., the pro-coagulants factor IX and factor VIII, and the
CC anti-coagulants protein C and antithrombin III), human
CC alpha-1-antitrypsin, PEA-3 protein and reporter proteins such as
CC luciferase. Preferred promoters for use in such age-regulatable
CC expression vectors include the human factor IX promoter, the T7 promoter,
CC the T3 promoter and the SP6 promoter. The expression vectors of the
CC invention may be used in gene therapy to provide age-related and/or
CC liver-specific expression of target genes. Age-regulatable constructs may
CC be used in the treatment of such age-related conditions such as
CC thrombosis, cardiovascular disease, diabetes, Alzheimer's disease,
CC Parkinson's disease, cancer, osteoporosis, osteoarthritis and dementia.
CC Specifically, they may be used to express factor IX antisense mRNA in the
CC treatment of thrombotic conditions associated with the natural
CC age-related rise in factor IX expression. Transgenic cells or animals
CC that contain vectors of the invention are useful as models of these
CC diseases, in screening for potential therapeutic agents and for studying

CC normal processes such as ageing and gene expression. Fragments and
CC homologues of age-related regulatory sequences, are useful as probes to
CC detect, isolate or identify other such sequences in samples. The present
CC sequence represents the hFIX gene.
XX
SQ Sequence 38059 BP; 12326 A; 7397 C; 7441 G; 10895 T; 0 other;

Query Match 1.3%; Score 21; DB 22; Length 38059;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 827 cagtggtcagtgcaatcataa 847
|||||
Db 17480 CAGTGGTCAGTGCAATCATAA 17460

RESULT 5
AAC02802
ID AAC02802 standard; cDNA; 390 BP.
XX
AC AAC02802;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 2800.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX Homo sapiens.
OS
XX EP1033401-A2.
PN
XX 06-SEP-2000.
PD
XX 21-FEB-2000; 2000EP-0200610.
PF
XX 26-FEB-1999; 99US-0122487.
PR
XX (GEST) GENSET.
PA
XX Dumas Milne Edwards J, Duclert A, Giordano J;
PI WPI; 2000-500381/45.
DR P-PSDB; AAG02796.
DR
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
PT
XX Claim 1; SEQ ID 2800; 71pp + CD-ROM; English.
PS
XX The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. An ORF has been identified within the
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC derived from 30 different tissues. EST sequences usually correspond
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC well suited for isolating cDNA sequences derived from the 5' ends of
CC mRNAs and even in those cases where longer cDNA sequences have been
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC mRNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC gene therapy and chromosome mapping procedures. They are used to obtain
CC upstream regulatory sequences and to design expression and secretion
CC vectors.
XX
SQ Sequence 390 BP; 78 A; 103 C; 131 G; 76 T; 2 other;

Query Match 1.3%; Score 20; DB 21; Length 390;
Best Local Similarity 100.0%; Pred. No. 11;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 505 gtattctcatctacgaggg 524
|||||
Db 289 gtattctcatctacgaggg 308
|||||

RESULT 6
AAT12188
ID AAT12188 standard; cDNA; 379 BP.
XX
AC AAT12188;
XX
DT 09-AUG-1996 (first entry)
XX
DE pJG4-5-CDK-BP cDNA clone #68 3' fragment.
XX
KW Cell cycle; regulation; G1 phase; proliferation; tumorigenesis;
KW cyclin dependent kinase; differentiation; CDK4 inhibitor; agonist;
KW antagonist; ss.
XX
OS Synthetic.
XX
PN WO9533819-A2.
XX
PD 14-DEC-1995.
XX
PF 02-JUN-1995; 95WO-US07113.
XX
PR 02-JUN-1994; 94US-0253155.
XX
PA (MITO-) MITOTIX INC.
XX
PI Draetta G, Gyuris J;
XX
DR WPI; 1996-040227/04.
XX
PT Cyclin-dependent kinase-4 binding protein - used in the isolation of
PT (ant)agonists of cell cycle regulation.
XX
PS Claim 30; Page 93; 115pp; English.
XX
CC AAT12186-T12207 are internal or 3' fragments of cDNA clones of the
CC plasmid pJG4-5-CDK-BP. The 5' end of the clones encode cyclin dependent
CC kinase 4 (CDK4) binding proteins (CDK4-BP), which may be used in an assay
CC for screening test compounds as inhibitors of CDK/CDK4-BP interaction.
CC The complexes formed by CDK4 and D-type cyclins are strongly implicated
CC in the control of the early G1 phase of the cell cycle and are strong
CC candidates for controlling and/or preventing tumorigenesis and the
CC onset of cancer. Nucleic acids encoding CDK4-BP or fragments of these
CC may be used as probes/primers to diagnose the presence or absence of
CC genetic lesions in a gene encoding 1 of 24 claimed CDK4-BP, and hence to
CC diagnose the risk for a subject of developing a cell-proliferation
CC associated disorder (e.g. cancer).
XX
SQ Sequence 379 BP; 105 A; 96 C; 80 G; 96 T; 2 other;

Query Match 1.2%; Score 19; DB 17; Length 379;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1027 aacgtgtacaagagactg 1045
|||||
Db 100 aacgtgtacaagagactg 118
|||||

RESULT 7
ABL18293
ID ABL18293 standard; DNA; 718 BP.
XX
AC ABL18293;
XX

DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 6352.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 6352; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 718 BP; 176 A; 172 C; 164 G; 206 T; 0 other;

Query Match 1.2%; Score 19; DB 23; Length 718;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 890 atggaattacagatctoga 908
|||||
Db 218 atggaattacagatctoga 236
|||||

RESULT 8
AAF63760
ID AAF63760 standard; DNA; 1170 BP.
XX
AC AAF63760;
XX
DT 03-APR-2001 (first entry)
XX
DE Drosophila gustatory receptor GR59D.1 DNA sequence.
XX
KW Gustatory receptor; fruit fly; taste; pheromone; semiochemical;
KW crop damage; pest control; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200077208-A2.
XX
PD 21-DEC-2000.
XX

PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0241826.
PR 08-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.

PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-483426/52.

Nucleic acids encoding human immune/hematopoietic antigen polypeptides, useful for preventing, diagnosing and/or treating cancers and metastasis -

Disclosure; SEQ ID NO 38402; 3071pp + Sequence Listing; English.

AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I) amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic activity, and can be used in gene therapy and vaccine production. (I) proteins and polynucleotides may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate (I) expression. For example, they may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of (I) by expressing inactive proteins or to supplement the patients own production of (I). Additionally, (I) polynucleotides may be used to produce the secreted (I), by inserting the nucleic acids into a host cell and culturing the cell to express the protein. (I) proteins and polynucleotides may be used to prevent, diagnose and treat immune/haematopoietic-related diseases, especially cancers and cancer metastases of haematopoietic-derived cells. AAK64703 to AAK87694 represent human immune/haematopoietic antigen genomic sequences from the present invention. AAK54942 to AAK54950 and AAM82169

CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 2302 BP; 401 A; 572 C; 712 G; 617 T; 0 other;

Query Match 1.2%; Score 19; DB 22; Length 2302;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1027 aacgctgtacaagagactg 1045
|||||
Db 2204 AACGCTGTACAGAGACTG 2186

RESULT 11
ABL18292/c
ID ABL18292 standard; DNA; 2718 BP.
XX
AC ABL18292;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 6349.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 6349; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
PS Claim 1; SEQ ID NO 6349; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2718 BP; 803 A; 546 C; 582 G; 787 T; 0 other;

Query Match 1.2%; Score 19; DB 23; Length 2718;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 890 atggaattacagatctcga 908
|||||
Db 1501 ATGGAATTACAGATCTCGA 1483

RESULT 12
ABL22926/c
ID ABL22926 standard; DNA; 4519 BP.
XX
AC ABL22926;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 20251.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 20251; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 4519 BP; 1283 A; 1005 C; 1023 G; 1208 T; 0 other;

Query Match 1.2%; Score 19; DB 23; Length 4519;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1267 gccctactctctaccaaa 1285
|||||
Db 3540 GCCCTACTCCTCTACCAA 3522

RESULT 13
ABL10880
ID ABL10880 standard; cDNA; 12332 BP.
XX
AC ABL10880;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 27122.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.


```
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX PA (PEKE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
DR WPI; 2001-656860/75.
DR P-PSDB; ABB66777.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX PS Claim 1; SEQ ID NO 27122; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 12332 BP; 3452 A; 2580 C; 2685 G; 3615 T; 0 other;

Query Match 1.2%; Score 19; DB 23; Length 12332;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 260 tctacatctgcgcgttt 278
Db 6219 tctacatctgcgcgttt 6237

RESULT 14
ABA19635
ID ABA19635 standard; DNA; 27869 BP.
XX AC ABA19635;
XX DT 23-JAN-2002 (first entry)
XX DE Human nervous system related polynucleotide SEQ ID NO 11966.
XX KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antisklking; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX OS Homo sapiens.
XX PN WO200159063-A2.
XX PD 16-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US01334.
XX PR 31-JAN-2000; 2000US-0179065.
XX PR 04-FEB-2000; 2000US-0180628.
XX PR 24-FEB-2000; 2000US-0184664.
XX PR 02-MAR-2000; 2000US-0186350.
XX PR 16-MAR-2000; 2000US-0189874.
XX PR 17-MAR-2000; 2000US-0190076.
XX PR 18-APR-2000; 2000US-0198123.
XX PR 19-MAY-2000; 2000US-0205515.
XX PR 07-JUN-2000; 2000US-0209467.
XX PR 28-JUN-2000; 2000US-0214886.
XX PR 30-JUN-2000; 2000US-0215135.
XX PR 07-JUL-2000; 2000US-0216647.
XX PR 07-JUL-2000; 2000US-0216880.
XX PR 11-JUL-2000; 2000US-0217487.
XX PR 11-JUL-2000; 2000US-0217496.
XX PR 14-JUL-2000; 2000US-0218290.
XX PR 26-JUL-2000; 2000US-0220963.
XX PR 26-JUL-2000; 2000US-0220964.
XX PR 14-AUG-2000; 2000US-0224518.
XX PR 14-AUG-2000; 2000US-0224519.
XX PR 14-AUG-2000; 2000US-0225213.
XX PR 14-AUG-2000; 2000US-0225214.
XX PR 14-AUG-2000; 2000US-0225266.
XX PR 14-AUG-2000; 2000US-0225267.
XX PR 14-AUG-2000; 2000US-0225268.
XX PR 14-AUG-2000; 2000US-0225270.
XX PR 14-AUG-2000; 2000US-0225447.
XX PR 14-AUG-2000; 2000US-0225757.
XX PR 14-AUG-2000; 2000US-0225758.
XX PR 14-AUG-2000; 2000US-0225759.
XX PR 18-AUG-2000; 2000US-0226279.
XX PR 22-AUG-2000; 2000US-0226581.
XX PR 22-AUG-2000; 2000US-0226586.
XX PR 22-AUG-2000; 2000US-0227182.
XX PR 23-AUG-2000; 2000US-0227009.
XX PR 30-AUG-2000; 2000US-0228924.
XX PR 01-SEP-2000; 2000US-0229287.
XX PR 01-SEP-2000; 2000US-0229343.
XX PR 01-SEP-2000; 2000US-0229344.
XX PR 01-SEP-2000; 2000US-0229345.
XX PR 05-SEP-2000; 2000US-0229509.
XX PR 05-SEP-2000; 2000US-0229513.
XX PR 06-SEP-2000; 2000US-0230437.
XX PR 06-SEP-2000; 2000US-0230438.
XX PR 08-SEP-2000; 2000US-0231242.
XX PR 08-SEP-2000; 2000US-0231243.
XX PR 08-SEP-2000; 2000US-0231244.
XX PR 08-SEP-2000; 2000US-0231413.
XX PR 08-SEP-2000; 2000US-0231414.
XX PR 08-SEP-2000; 2000US-0232080.
XX PR 08-SEP-2000; 2000US-0232081.
XX PR 12-SEP-2000; 2000US-0231968.
XX PR 14-SEP-2000; 2000US-0232397.
XX PR 14-SEP-2000; 2000US-0232398.
XX PR 14-SEP-2000; 2000US-0232399.
XX PR 14-SEP-2000; 2000US-0232400.
XX PR 14-SEP-2000; 2000US-0232401.
XX PR 14-SEP-2000; 2000US-0233063.
XX PR 14-SEP-2000; 2000US-0233064.
XX PR 14-SEP-2000; 2000US-0233065.
XX PR 21-SEP-2000; 2000US-0234223.
XX PR 21-SEP-2000; 2000US-0234274.
XX PR 25-SEP-2000; 2000US-0234997.
XX PR 25-SEP-2000; 2000US-0234998.
XX PR 26-SEP-2000; 2000US-0235484.
XX PR 27-SEP-2000; 2000US-0235834.
XX PR 27-SEP-2000; 2000US-0235836.
XX PR 29-SEP-2000; 2000US-0236327.
XX PR 29-SEP-2000; 2000US-0236367.
XX PR 29-SEP-2000; 2000US-0236368.
XX PR 29-SEP-2000; 2000US-0236369.
```

PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-541565/60.
XX
PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,

PT useful for preventing, diagnosing and/or treating nervous system
PT cancers and metastases -
XX
PS Disclosure; SEQ ID NO 11966; 1701pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (AB11004-ABA21534) and proteins
CC (AB14678-AB18001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (antagonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 27869 BP; 7054 A; 5673 C; 6322 G; 7820 T; 0 other;

Query Match 1.2%; Score 19; DB 22; Length 27869;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 155 ttgtgatttgcaacagttg 173
Db 26107 ttgtgatttgcaacagttg 26125
|||||

RESULT 15
AAK66517/c
ID AAK66517 standard; DNA; 27869 BP.
XX
AC AAK66517;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human Immune/haematopoietic antigen genomic sequence SEQ ID NO:21329.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PE 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.

PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-483426/52.
DR
XX
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 21329; 3071pp + Sequence Listing; English.
XX
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially

Search completed: April 27, 2002, 05:29:34
Job time: 13857 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2002, 05:34:37 ; Search time 7820.58 Seconds
(without alignments)
4222.459 Million cell updates/sec

Title: US-09-882-694A-7
Perfect score: 1578
Sequence: 1 atggactccagaccagtg.....aattccaagttgggcacatga 1578

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1797656 seqs, 10463268293 residues

Word size : 0
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl:*

- 1: gb_ba:*
- 2: gb_htg:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vi:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_sts:*
- 28: em_un:*
- 29: em_vi:*
- 30: em_htg_hum:*
- 31: em_htg_inv:*
- 32: em_htg_other:*
- 33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	DB Length	ID	Description

1	937	59.4	1578	6	AX076848	Sequence
2	937	59.4	1764	6	AX076847	Sequence
C 3	22	1.4	38209	9	AP000292	Homo sapi
C 4	22	1.4	40697	9	AF039905	Homo sapi
C 5	22	1.4	100000	9	AP000043	Homo sapi
C 6	22	1.4	100000	9	AP000111	Homo sapi
C 7	22	1.4	100000	9	AP000187	Homo sapi
C 8	22	1.4	340000	9	AP001716	Homo sapi
C 9	21	1.3	11873	6	HSFAC1XG	H.sapiens f
C 10	21	1.3	38059	9	HUMFIXG	Human coagu
C 11	21	1.3	153615	2	AL645665	Homo sapi
C 12	21	1.3	158557	9	HS88D7	Human DNA
13	20	1.3	2154	8	AF052582	Fusarium
14	20	1.3	2770	10	RNPFKL	X58865 Rat PFK-L m
15	20	1.3	2879	9	BC007536	Homo sapi
16	20	1.3	2914	9	HSPEKLA	Human liver
17	20	1.3	2919	9	BC008964	Homo sapi
18	20	1.3	2920	9	BC009919	Homo sapi
19	20	1.3	3385	9	BC006422	Homo sapi
20	20	1.3	14023	1	AE007062	Mycobacte
C 21	20	1.3	46778	9	AL356003	Human DNA
22	20	1.3	53450	1	MTV018	Human DNA
C 23	20	1.3	81908	9	AL590623	Human DNA
24	20	1.3	83408	8	AB078516	Arabidops
C 25	20	1.3	85382	9	AL390793	Human DNA
C 26	20	1.3	115890	2	AP002824	Homo sapi
27	20	1.3	153094	2	AC091973	Homo sapi
28	20	1.3	153819	2	AC109466	Homo sapi
29	20	1.3	157078	9	AC096727	Homo sapi
30	20	1.3	179415	2	AC024286	Homo sapi
C 31	20	1.3	188872	9	AC020760	Homo sapi
32	20	1.3	193910	2	AC040996	Homo sapi
33	20	1.3	232506	2	AC109822	Homo sapi
34	19	1.2	177	9	HSPFK02	Human PFKL
35	19	1.2	379	6	I76223	Sequence 51
36	19	1.2	955	5	TCU94815	Trogon curu
37	19	1.2	955	5	TVU94816	Trogon viol
C 38	19	1.2	1155	9	BC001822	Homo sapi
39	19	1.2	1170	6	AX058757	Sequence
C 40	19	1.2	1256	9	HSU40343	Human CDK i
41	19	1.2	1901	3	AY071767	Drosophil
42	19	1.2	3470	3	AY052145	Drosophil
C 43	19	1.2	6881	1	FPLTRAD	Plasmid F (
44	19	1.2	15847	9	AP001061	Homo sapi
C 45	19	1.2	16928	2	AC109117	Rattus no

ALIGNMENTS

RESULT	1					
AX076848						
LOCUS	AX076848	1578 bp	DNA	linear	PAT 22-FEB-2001	
DEFINITION	Sequence 7 from Patent WO0105980.					
ACCESSION	AX076848					
VERSION	AX076848.1	GI:13121521				
KEYWORDS						
SOURCE	Exophiala spinifera.					
ORGANISM	Exophiala spinifera					
	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Chaetothyriomycetes;					
	Chaetothyriales; Herpotrichiellaceae; mitosporic					
	Herpotrichiellaceae; Exophiala.					
REFERENCE	1 {bases 1 to 1578}					
AUTHORS	Duvick,J.P., Maddox,J., Gilliam,J., Folkerts,O. and Crasta,O.R.					
TITLE	Compositions and methods for fumonisins detoxification					
JOURNAL	Patent: WO 0105980-A 7 25-JAN-2001;					
	Pioneer Hi-Bred International Inc. (US) ; Curagen Corporation (US)					
FEATURES	Location/Qualifiers					
Source	1..1578					
	/organism="Exophiala spinifera"					
	/db_xref="taxon:91928"					
	/note="permease, coding sequence"					
BASE COUNT	324 a	446 c	385 g	423 t		

ORIGIN									
Query Match 59.4%; Score 937; DB 6; Length 1578;									
Best Local Similarity 99.2%; Pred. No. 0;									
Matches 1537; Conservative 0; Mismatches 12; Indels 0; Gaps 0;									
Qy	30	gaaagcgggagagcagacaaacgaagaacacagagacgagcggcgggcgaggtggtgcgctc	89						
Db	30	gaaagcgggagagcagacaaacgaagaacacagagacgagcggcgggcgaggtggtgcgctc	89						
Qy	90	cgaagtcctgaacgttcctctggaagaagaacaatttggcaccatcaccatcggtgcctt	149						
Db	90	cgaagtcctgaacgttcctctggaagaagaacaatttggcaccatcaccatcggtgcctt	149						
Qy	150	ggcctttgtgatttgcaacagttgggtggtatctcagcgagtcctccagctcgccctact	209						
Db	150	ggcctttgtgatttgcaacagttgggtggtatctcagcgagtcctccagctcgccctact	209						
Qy	210	agcgggggggccctcactctcttaccggcalccatcaatcagtcactctcgtctacatctg	269						
Db	210	agcgggggggccctcactctcttaccggcalccatcaatcagtcactctcgtctacatctg	269						
Qy	270	catcgctttctcattagccgaactgaccagcgctctacccgactgcggtggtgccaatatca	329						
Db	270	catcgctttctcattagccgaactgaccagcgctctacccgactgcggtggtgccaatatca	329						
Qy	330	ttttcgctcgatcctcctggcaccaaaatcaatcaatcgagcatttcatacgtgctggact	389						
Db	330	ttttcgctcgatcctcctggcaccaaaatcaatcaatcgagcatttcatacgtgctggact	389						
Qy	390	cgtgctgctgcttctcattggtatcgctatcggaagctcagtgaccatgatascgtcaca	449						
Db	390	cgtgctgctgcttctcattggtatcgctatcggaagctcagtgaccatgatascgtcaca	449						
Qy	450	gabcggcgctgtagccgcttatagtcacacatactccaggattcgtggcatgctt	509						
Db	450	gabcggcgctgtagccgcttatagtcacacatactccaggattcgtggcatgctt	509						
Qy	510	cctcatctacgagggagtcgcgctggtgctctgttgcacttgcacttgcctgaaaag	569						
Db	510	cctcatctacgagggagtcgcgctggtgctctgttgcacttgcacttgcctgaaaag	569						
Qy	570	aaaccccttggttcattgaaatcggaatcgccctcagcagtcgctctctcgtgatctcctt	629						
Db	570	aaaccccttggttcattgaaatcggaatcgccctcagcagtcgctctctcgtgatctcctt	629						
Qy	630	tatcgccattctagcgcggtccaaaccccaaggctccaaactcacaggatggtgactgctt	689						
Db	630	tatcgccattctagcgcggtccaaaccccaaggctccaaactcacaggatggtgactgctt	689						
Qy	690	gagcaactatactggtggtccgagggcgtctgcttccatcctggtggtccttcgacatcctg	749						
Db	690	gagcaactatactggtggtccgagggcgtctgcttccatcctggtggtccttcgacatcctg	749						
Qy	750	cttcatgttcatggtggtgagcagcaatgcatctggtggtgagaaatgacagatgctgc	809						
Db	750	cttcatgttcatggtggtgagcagcaatgcatctggtggtgagaaatgacagatgctgc	809						
Qy	810	tcgtacgggtaccacaaagcagtggtcagtgcaatcataatggcttctgaccgcctttcc	869						
Db	810	tcgtacgggtaccacaaagcagtggtcagtgcaatcataatggcttctgaccgcctttcc	869						
Qy	870	atatacaatcgagttctgtatggaattacagatctcgactctattctaaagtccgcgg	929						
Db	870	atatacaatcgagttctgtatggaattacagatctcgactctattctaaagtccgcgg	929						
Qy	930	ctatatctcattcgagacaatgacgcagtccttcggtcgctcagttttgcaacggtcct	989						
Db	930	ctatatctcattcgagacaatgacgcagtccttcggtcgctcagttttgcaacggtcct	989						
Qy	990	cctcatggtggtgctatcgtgattggtcctctctcgcctcaacgctgtacagagactgcgc	1049						

Db	990	CTCATGTGCGGTATCGTGTATGCGCCTTCTTTCGCCCTCAACGGCTGTACAGAGACTGCGTC	1049
Qy	1050	tcgactcactggagctttgcccgggacaatgggtggtattttccactcatctcgaacg	1109
Db	1050	TCGACTACCTGGAGCTTTGCCGGGACAATGGCTGCTATTTTCCACTCATCTCGAAGC	1109
Qy	1110	cattcatcccgctggcaagttcctgtttggtctctattcgcgacctggggaattctggc	1169
Db	1110	CATTCACTCCGCTGGCAAGTCTCTGTTTGGTCTCTATTTCGCGACCTGGGGAATTCTGGC	1169
Qy	1170	cacatgcgagtgatattttctaggttctagcacagctttcaatgccttggtcaattccgc	1229
Db	1170	CACATGCGGATGTATATTCTAGGTTCTAGCACAGCTTTCAATGCCCTTGGTCAATTCCGC	1229
Qy	1230	cgttgtactccagcaactctctctctgatcccaatcgccctactctctaccacaaagcg	1289
Db	1230	CGTTGTACTCCAGCAACTCTCTTCTTGATCCCAATCGCCTACTCTCTACCAAAAGCG	1289
Qy	1290	agatccaaagtctctgcccagcactcgtcttttggttaccgcgctgggaatcggtttct	1349
Db	1290	AGATCCAAAGTTCTTGCCGAGCACTCGTGCTTTTGTGTTACCGCGTGAATCGGGTTCT	1349
Qy	1350	ggtcaatgtcgtagcgggtggtcttcacgctccgctcaccactggtttttcagcttccact	1409
Db	1350	GGTCAATGTCTAGCGGTGGTCTTTCACGTCCGTCACCACTGTGTTTTTTCAGCTTCCACT	1409
Qy	1410	gacctgcctacggccgctcaacccatgaattacacaagtgcgattatagcggttcact	1469
Db	1410	GACCGTGCTACGGCCGCGTCAACCATGAATTACACAAGTGGATATATAGCGCTTGCAC	1469
Qy	1470	tgctcttggtgtcttgaaactgggtggtcgtgcagtcaggaagcattatcagggacccactt	1529
Db	1470	TGCTCTTGGTGCTTGAACCTGGTCTGTCATGCCAGGAAGCATTTATCAGGGACCCACTT	1529
Qy	1530	ggagcttgacggacgggtcgtcggagcagaatttcaagttgggcccata	1578
Db	1530	GGAGCTTGACGGACGGGTCTGCGGAGCAGAAATTCAAGTTGGGCCCATGA	1578
RESULT 2			
LOCUS	AX076847	Sequence 6 from Patent WO0105980.	1764 bp DNA linear PAT 22-FEB-2001
DEFINITION	AX076847		
ACCESSION	AX076847		
VERSION	AX076847.1	GI:13121520	
KEYWORDS	Exophiala spinifera.		
SOURCE	Exophiala spinifera.		
ORGANISM	Exophiala spinifera		
	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Chaetothyriomycetes;		
	Chaetothyriales; Herpotrichiellaceae; mitosporic		
	Herpotrichiellaceae; Exophiala.		
REFERENCE	1 (bases 1 to 1764)		
AUTHORS	Duvick, J.P., Maddox, J., Gilliam, J., Folkerts, O. and Crasta, O.R.		
TITLE	Compositions and methods for fumonisins detoxification		
JOURNAL	Patent: WO 0105980-A 6 25-JAN-2001;		
	Pioneer Hi-Bred International Inc. (US) ; Curagen Corporation (US)		
FEATURES	Location/Qualifiers		
source	1..1764		
	/organism="Exophiala spinifera"		
	/db_xref="taxon:91928"		
	/note="permease"		
BASE COUNT	397 a 478 c 432 g 454 t	3 others	
ORIGIN			
Query Match 59.4%; Score 937; DB 6; Length 1764;			
Best Local Similarity 99.2%; Pred. No. 0;			
Matches 1537; Conservative 0; Mismatches 12; Indels 0; Gaps 0;			
Qy	30	gaaagcgggagagcagacaaacgaagaacacagagacgagcggcgggcgaggtggtgcgctc	89
Db	30	gaaagcgggagagcagacaaacgaagaacacagagacgagcggcgggcgaggtggtgcgctc	89
Qy	90	cgaagtcctgaacgttcctctggaagaagaacaatttggcaccatcaccatcggtgcctt	149
Db	90	cgaagtcctgaacgttcctctggaagaagaacaatttggcaccatcaccatcggtgcctt	149
Qy	150	ggcctttgtgatttgcaacagttgggtggtatctcagcgagtcctccagctcgccctact	209
Db	150	ggcctttgtgatttgcaacagttgggtggtatctcagcgagtcctccagctcgccctact	209
Qy	210	agcgggggggccctcactctcttaccggcalccatcaatcagtcactctcgtctacatctg	269
Db	210	agcgggggggccctcactctcttaccggcalccatcaatcagtcactctcgtctacatctg	269
Qy	270	catcgctttctcattagccgaactgaccagcgctctacccgactgcggtggtgccaatatca	329
Db	270	catcgctttctcattagccgaactgaccagcgctctacccgactgcggtggtgccaatatca	329
Qy	330	ttttcgctcgatcctcctggcaccaaaatcaatcaatcgagcatttcatacgtgctggact	389
Db	330	ttttcgctcgatcctcctggcaccaaaatcaatcaatcgagcatttcatacgtgctggact	389
Qy	390	cgtgctgctgcttctcattggtatcgctatcggaagctcagtgaccatgatascgtcaca	449
Db	390	cgtgctgctgcttctcattggtatcgctatcggaagctcagtgaccatgatascgtcaca	449
Qy	450	gabcggcgctgtagccgcttatagtcacacatactccaggattcgtggcatgctt	509
Db	450	gabcggcgctgtagccgcttatagtcacacatactccaggattcgtggcatgctt	509
Qy	510	cctcatctacgagggagtcgcgctggtgctctgttgcacttgcacttgcctgaaaag	569
Db	510	cctcatctacgagggagtcgcgctggtgctctgttgcacttgcacttgcctgaaaag	569
Qy	570	aaaccccttggttcattgaaatcggaatcgccctcagcagtcgctctctcgtgatctcctt	629
Db	570	aaaccccttggttcattgaaatcggaatcgccctcagcagtcgctctctcgtgatctcctt	629
Qy	630	tatcgccattctagcgcggtccaaaccccaaggctccaaactcacaggatggtgactgctt	689
Db	630	tatcgccattctagcgcggtccaaaccccaaggctccaaactcacaggatggtgactgctt	689
Qy	690	gagcaactatactggtggtccgagggcgtctgcttccatcctggtggtccttcgacatcctg	749
Db	690	gagcaactatactggtggtccgagggcgtctgcttccatcctggtggtccttcgacatcctg	749
Qy	750	cttcatgttcatggtggtgagcagcaatgcatctggtggtgagaaatgacagatgctgc	809
Db	750	cttcatgttcatggtggtgagcagcaatgcatctggtggtgagaaatgacagatgctgc	809
Qy	810	tcgtacgggtaccacaaagcagtggtcagtgcaatcataatggcttctgaccgcctttcc	869
Db	810	tcgtacgggtaccacaaagcagtggtcagtgcaatcataatggcttctgaccgcctttcc	869
Qy	870	atatacaatcgagttctgtatggaattacagatctcgactctattctaaagtccgcgg	929
Db	870	atatacaatcgagttctgtatggaattacagatctcgactctattctaaagtccgcgg	929
Qy	930	ctatatctcattcgagacaatgacgcagtccttcggtcgctcagttttgcaacggtcct	989
Db	930	ctatatctcattcgagacaatgacgcagtccttcggtcgctcagttttgcaacggtcct	989
Qy	990	cctcatggtggtgctatcgtgattggtcctctctcgcctcaacgctgtacagagactgcgc	1049

QY 90 cgagtcctgaagcttctctgtgagagaacaatttggcaccatcaccatcggtgctt 149
Db 94 CGAGTCCCTGAACGTTCCCTCTGGAGAGAAACAATTTGGACCATCACCATCGTGCCTT 153
QY 150 ggcctttgtgattgcaacagtttgggtggtatctcagcagagtcctcagctcgccctact 209
Db 154 GGCCTTTGTGATTGCAACAGTTGGGTGGTATCTCAGGCAGTCTCCAGCTCGCCCTACT 213
QY 210 agcggggggcccgctcactctcctttacggcgcaccccaatcagtcactctcgtctacatctg 269
Db 214 AGCGGGGGGGCCCTCACTCTCCCTTACGGCATCTTAATCAGTACTCTCGTCTACATCTG 273
QY 270 catcgctttctcattagccgaactgaccagcgtctacccgactgccgggtggcccaatca 329
Db 274 CATCGCTTCTCATAGCCGAACCTGACGAGGCTTACCCGACTGCCGGTGGCCAAATATCA 333
QY 330 ttttgcgtcgatccttggcaccaaaatacaatcaatcgagcatctcattcatcgtgtgcgact 389
Db 334 TTTTGGCTCGATCCTGGCACCAAAATCAATCAATCGGAGCATTCATACGTTGCGGACT 393
QY 390 cgtgctgttgccttcacatggatcgctatcggaagctcagtgaccatgatacctgctcaaca 449
Db 394 CGTGCTGTTGCTTTTCATGGATCGCTATCGGAAGCTCAGTGACCATGATACCTGCTCAACA 453
QY 450 gatcccggtcgctgataccgcttatagtcacacatactcccaggatctgtggcatgtt 509
Db 454 GATCCCGGCGCTGATAGCCGCTATAGTCACACATACTCCCAGGATTCGTGGCATGTCTT 513
QY 510 cctcatctacgaggagtcgcgctggtggtgctcttctgttcaacttgttgccttgaagaag 569
Db 514 CCTCATCTACGAGGAGTCCGCTGGTGGTGTCTTGTTCAACTGTTTGGCCCTGAAAAG 573
QY 570 aaacccttgggttcatgaaatcggaattcgccctcacgatacgctctctcgtgatctcctt 629
Db 574 AAACCTTGGGTTTCATGAATCGGATTCGGCCTCACGATCGCTCTCTTCGTGATCTCCTT 633
QY 630 tatcgccattctagcgggtccaaccccccaaggctccaaactcacaggtatggactgcttg 689
Db 634 TATCGCCATTCTAGCCGGTCCAACCCCCAAGGCTCCAAACTCACAGGTATGGACTGCTTG 693
QY 690 gagcaactatactggtgtccgacggtgctgcttctcactcctggccttctcgacatcctg 749
Db 694 GACCAACTATACTGGCTGGTCCGACGGGCTGTCTTCTCATCTCTGGSCCTTTCGACATCCTG 753
QY 750 ctctcatgttcattggttggacgcagcaatgcatctggtggtgaaagatgcacagatgctgc 809
Db 754 CTTCATGTTTCATTGGCTTGGACGCAGCAATGCATCTGGCTGAAGATGCACAGATGCTGC 813
QY 810 tegtacggtacccaaagcagtggtcagtgcaatcataaattggcttctgcaccgcctttcc 869
Db 814 TCGTACGGTACCCAAACGAGTGGTTCAGTGCAATCATATAATGGCTTCTGCACCGCCTTCC 873
QY 870 atatacaatcgagttctgtatggaattacagatctcgactctatttctaagttccgcgg 929
Db 874 ATATACAATCGCAGTTCTGTATGGAATTACAGATCTCGACTCTATTTCTAAGTTCCGCCGG 933
QY 930 ctatatccattcagagacaatgacgcagtccttctggtcgctcagtttttgcaacggtcct 989
Db 934 CTATATTCCATTTCGAGACAATGAGGCAGTCTGCTCGAATTCGGCTTTTTCACACGGTCTCT 993
QY 990 ctaatgtggcgggtatcgtgatggccttcttgcgcctcaacgctgtacaagagactgcgtc 1049
Db 994 CTATGTGGCGGTATCGTGATGGCTTCTTTCGCCCTCAACGCTGTACAGAGACTGCGTC 1053
QY 1050 tgcactcacctggagcttttggccgggacaattgggtggtattttccactcatctcgaacg 1109
Db 1054 TCGACTCACCTGGAGCTTTGCCCGGGACAATGGGCTGGTATTTTCCACTCATCTCGAACG 1113
QY 1110 catcattccccgctggcaagttcctgtttgtctctattcgcgacctgggggaattctggc 1169
Db 1114 CATTCATCCCCGCTGGCAAGTTCCTGTTTGGTCTCTATTTCGGGACCTGGGGAAATCTGGC 1173
QY 1170 cacatgcgatgtatatatttctaggttctctagcacagcttttcaatgccttggtcattccgc 1229

Db 1174 CACATGGGATGTATATTCTAGGTTCTAGCACAGCTTTCATGCTTGGTCAATCCGC 1233
QY 1230 cgttgtactccagcaactctccttctctgatcccaatcgccctactctctacaaaagcg 1289
Db 1234 CGTTGTACTCCAGCAACTCTCTCTCTGATCCCAATCGCCCTACTCTCTACCAAAAGCG 1293
QY 1290 agatccaaagtcttgcgagcactcgtgtcttttgtgtaccgcgtggaatcggtttct 1349
Db 1294 AGATCCAAAGTCTTTCGGAGCACTCGTGTCTTTGTGTTCACCGCTGGAATCGGGTTCT 1353
QY 1350 ggtcaatgtgctagcgggtggttcttcaacgctccgctcaacactgtgttttccagcttccact 1409
Db 1354 GGTCATGTGCTAGCGGTGGTCTTCACGTCCTCACCACCTGTGTTTTCAGCTTCCCACT 1413
QY 1410 gaccgtgcctacggcgcgtcaaccatgaattacacaagtcgattataggcgttgcaact 1469
Db 1414 GACCGTGCTTACGGCCGCTCAACCATGAATTACACAAGTCCGATTATAGGCGTTGCAC 1473
QY 1470 tgccttgggtgttgaactgggtcgtgcatgccagggaagcattatcagggacccactt 1529
Db 1474 TGCTCTTGGTGTCTTGAACCTGGTCTGCATGCCAGGAGCATATATCAGGGACCCACTT 1533
QY 1530 ggagcttgacggacgggtcgtcggagcagaatttcaagttgggccaatga 1578
Db 1534 GGAGCTTGACGACGGGTCTGTCGGAGCAGAAATTTCAAGTTGGGCCATGA 1582

RESULT 3
AP000292/c
LOCUS
DEFINITION
AP000292
Homo sapiens genomic DNA, chromosome 21q22.1, D21S226-AML region,
clone:Q95D4, complete sequence.

ACCESSION
AP000292
VERSION
AP000292.1 GI:4835661
KEYWORDS
HTG.

SOURCE
Homo sapiens DNA, clone:Q95D4.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
AUTHORS
1 (bases 1 to 38209)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y. and Sakaki,Y.

TITLE
JOURNAL
REFERENCE
AUTHORS
Homo sapiens 38,209bp genomic DNA of 21q22.1
Published Only in DataBase (1999) In press
2 (bases 1 to 38209)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y. and Sakaki,Y.

TITLE
JOURNAL
Direct Submission
Submitted (13-MAY-1999) to the DDBJ/EMBL/GenBank databases.
Masahira Hattori, The Institute of Physical and Chemical Research
(RIKEN), Genomic Sciences Center (GSC); Kitasato Univ., 1-15-1
Kitasato, Sagamihara, Kanagawa 228-8555, Japan
(E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-42-778-9923, Fax:81-42-778-9924)

COMMENT
The sequence is a part of the data (ACCESSION No. AP000174 -
AP000194).

FEATURES
source
1. .38209
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/clone="Q95D4"
/map="21q22.1"

BASE COUNT
ORIGIN
10658 a 8601 c 7890 g 11060 t

Query Match 1.4%; Score 22; DB 9; Length 38209;
Best Local Similarity 100.0%; Pred. No. 1;

/evidence-not_experimental
12828. .12963
/rpt_family="L1M4"
/evidence-not_experimental
13020. .13319
/rpt_family="AluY"
/evidence-not_experimental
13154. .13295
/note="GC score = 6.80 (142bp)"
/note="Region: GC content"
/evidence-not_experimental
13368. .13480
/rpt_family="L1ME"
/evidence-not_experimental
complement(13487. .13851)
/rpt_family="MLT1E"
/evidence-not_experimental
13945. .14002
/note="homology = 94.80%, score = 23, counts = 2"
/rpt_type=tandem
/rpt_unit-atgaaggcattaattatgtctgcaaac
/evidence-not_experimental
14514. .14813
/rpt_family="AluJb"
/evidence-not_experimental
15160. .15257
/rpt_family="FLAM_C"
/evidence-not_experimental
15320. .15635
/rpt_family="L2"
/evidence-not_experimental
complement(15716. .15811)
/note="Xpound exon prediction, score = 76% (0%)"
/evidence-not_experimental
complement(15906. .16328)
/rpt_family="MLT2FB"
/evidence-not_experimental
16333. .16632
/rpt_family="L2"
/evidence-not_experimental
complement(16858. .17200)
/rpt_family="MLT1A2"
/evidence-not_experimental
complement(17346. .17473)
/rpt_family="MER20"
/evidence-not_experimental
complement(17709. .18006)
/rpt_family="AluSx"
/evidence-not_experimental
complement(19099. .19219)
/rpt_family="MER46"
/evidence-not_experimental
complement(19330. .19431)
/note="GRAIL, score = 42.000%, comment = marginal"
/evidence-not_experimental
complement(19809. .20415)
/rpt_family="MER4A"
/evidence-not_experimental
19969. .20066
/note="MZEFF, score = 50.8%"
/evidence-not_experimental
complement(20416. .20828)
/rpt_family="MER57B"

Query Match 1.4%; Score 22; DB 9; Length 40697;
Best Local Similarity 100.0%; Pred.No. 1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1177 ggatgtatatattctaggttcta 1198
|||||
Db 12042 GGATGTATATTTCTAGGTTCTA 12021

RESULT 5
AP000043/c
LOCUS
DEFINITION
AP000043 Homo sapiens genomic DNA, chromosome 21q22.1, segment 14/28, complete sequence.
ACCESSION
AP000043 AP000043.1 GI:3132353
VERSION
KEYWORDS
HTG.
SOURCE
Homo sapiens DNA, clone:TI866-f90A8f_7.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 100000)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Shiba,T. and Sakaki,Y.
TITLE Homo sapiens genomic DNA, chromosome 21q
JOURNAL Published Only in DataBase (1998) In press
REFERENCE
2 (bases 1 to 100000)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Shiba,T. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (11-MAY-1998) to the DDBJ/EMBL/GenBank databases.

Masahira Hattori, Kitasato University, Department of Science, JST Sequencing Laboratory; Kitasato 1-15-1, Sagamihara 228, Japan (E-mail:hattori@hgc.ims.u-tokyo.ac.jp, Tel:0427-78-9732, Fax:0427-78-9561)
This sequence is conducted by Kitasato University JST sequencing Laboratory as a JST sequencing team.
Principal Investigator:Yoshiyuki Sakaki Ph.D.
Phone: +81-3-5449-5622, Fax: +81-3-5449-5445, sakaki@hgc.ims.u-tokyo.ac.jp
Sub-leader: Tadayoshi Shiba Ph.D., Masahira Hattori Ph.D. The sequence is submitted by:Human Genome Sequencing in ALIS project of JST

Japan Science and Technology Corporation (JST)
5-3, Yonbancho, Chiyoda-ku, Tokyo 102-0028 Japan
For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (<http://www-alis.tokyo.jst.go.jp/HGS/top.html>) or send email to webmaster@www-alis.tokyo.jst.go.jp;

FEATURES
source

1. .100000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/clone="TI866-f90A8f_7"
/map="21q22.1"

BASE COUNT 27784 a 22015 c 21661 g 28540 t
ORIGIN

Query Match 1.4%; Score 22; DB 9; Length 100000;
Best Local Similarity 100.0%; Pred.No. 0.91;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1177 ggatgtatatattctaggttcta 1198
|||||
Db 30638 GGATGTATATTTCTAGGTTCTA 30617

RESULT 6
AP000111/c

LOCUS
DEFINITION
AP000111 Homo sapiens genomic DNA of 21q22.1, GART and AML related, Q78C10-149C3 region, segment 14/20, complete sequence.
ACCESSION
AP000111 AP000111.1 GI:4730845
VERSION
KEYWORDS
HTG.
SOURCE
Homo sapiens DNA.

ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (sites)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Shiba,T. and Sakaki,Y.
TITLE Homo sapiens 2,051,516bp genomic DNA of 21q22.1 GART and AML region

JOURNAL Published Only in DataBase (1999) In press
REFERENCE 2 (bases 1 to 100000)
AUTHORS Hirakawa,M., Yamaguchi,H., Imai,K. and Shimada,J.
TITLE Direct Submission
JOURNAL Submitted (15-APR-1999) to the DDBJ/EMBL/GenBank databases. Mika
Hirakawa, Japan Science and Technology Corporation (JST), Advanced
Databases Department; 5-3, Yonbancho, Chiyoda-ku, Tokyo 102-0081,
Japan (E-mail:mika@tokyo.jst.go.jp,
URL:http://www-alis.tokyo.jst.go.jp/, Tel:81-3-5214-8491,
Fax:81-3-5214-8470)
COMMENT This sequence is conducted by Kitasato University JST sequencing
Laboratory as a JST sequencing team.
Principal Investigator:Yoshiyuki Sakaki Ph.D.
Phone: +81-3-5449-5622, Fax : +81-3-5449-5445,
sakaki@hgc.ims.u-tokyo.ac.jp
Sub-leader: Tadayoshi Shiba Ph.D., Masahira Hattori Ph.D. The
sequence is submitted by Human Genome Sequencing in ALIS project of
JST.
Japan Science and Technology Corporation (JST)
5-3, Yonbancho, Chiyoda-ku, Tokyo 102-0081 Japan
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://www-alis.tokyo.jst.go.jp/HGS/) or send
email to webmaster@www-alis.tokyo.jst.go.jp.
FEATURES
source
Location/Qualifiers
1. .100000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.1"
66264..66403
/note="A006D32;The location is between each flanking site
of PCR primers."
/db_xref="GDB:4571163"
66282..66515
/note="SHGC-52112;The location is between each flanking
site of PCR primers."
/db_xref="GDB:6464755"
92086..92206
/note="SHGC-52139;The location is between each flanking
site of PCR primers."
/db_xref="GDB:6464581"
92099..92279
/note="RH18027;The location is between each flanking site
of PCR primers."
/db_xref="GDB:4572813"
BASE COUNT 27992 a 22056 c 21621 g 28331 t
ORIGIN
Query Match 1.4%; Score 22; DB 9; Length 100000;
Best Local Similarity 100.0%; Pred. No. 0.91;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1177 ggatgtatatttctaggttcta 1198
|||||
Db 36614 GGATGTATATTCTAGGTTCTA 36593
RESULT 7
AP000187/c
LOCUS AP000187 100000 bp DNA linear PRI 08-JAN-2000
DEFINITION Homo sapiens genomic DNA, chromosome 21q22.1, D21S226-AML region,
clone Q78C10-f32E9, segment 14/21, complete sequence.
ACCESSION AP000187
VERSION AP000187.1 GI:4827086
KEYWORDS HTG.
SOURCE Homo sapiens DNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 100000)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,

Fujiyama,A., Yada,T., Totoki,Y. and Sakaki,Y.
Homo sapiens 2,083,744bp genomic DNA of 21q22.1 (REGION:
D21S226-AML CLONE RANGE: Q78C10-f32E9)
Published Only in DataBase (1999) In press
2 (bases 1 to 100000)
JOURNAL Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (10-MAY-1999) to the DDBJ/EMBL/GenBank databases.
Masahira Hattori, The Institute of Physical and Chemical Research
(RIKEN), Genomic Sciences Center (GSC); Kitasato Univ., 1-15-1
Kitasato, Sagamihara, Kanagawa 228-8555, Japan
(E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-42-778-9923, Fax:81-42-778-9924)
COMMENT E. coli transposon insertion:The present data does not contain E.
coli transposon sequences which integrated in the
original/previous sequences. We determined the boundary between
the insertion and genomic sequences experimentally, removed the
insertion sequences, reconstituted the present data. The sequencing
project is supported by Japan Science Technology Corporation (JST)
and The Institute of Physical and Chemical Research (RIKEN).
FEATURES
source
Location/Qualifiers
1. .100000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.1"
BASE COUNT 27992 a 22056 c 21621 g 28331 t
ORIGIN
Query Match 1.4%; Score 22; DB 9; Length 100000;
Best Local Similarity 100.0%; Pred. No. 0.91;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1177 ggatgtatatttctaggttcta 1198
|||||
Db 36614 GGATGTATATTCTAGGTTCTA 36593
RESULT 8
AP001716/c
LOCUS AP001716 340000 bp DNA linear PRI 30-MAY-2000
DEFINITION Homo sapiens genomic DNA, chromosome 21q, section 60/105.
ACCESSION AP001716 AL163261 BA000005
VERSION AP001716.1 GI:7768717
KEYWORDS
SOURCE Homo sapiens DNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E.,
Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
Patterson,D., Reichwald,K., Rump,A., Schillhabel,M. and Schudy,A.
The DNA sequence of human chromosome 21. The chromosome 21 mapping
and sequencing consortium
Nature 405 (6784), 311-319 (2000)
JOURNAL Nature 405 (6784), 311-319 (2000)
MEDLINE 20289799
REFERENCE 2 (bases 1 to 340000)
AUTHORS Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E.,
Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
Patterson,D., Reichwald,K., Rump,A., Schillhabel,M., Schudy,A.,
Zimmermann,W., Rosenthal,A., Kudoh,J., Shibuya,K., Kawasaki,K.,
Asakawa,S., Shintani,A., Sasaki,T., Nagamine,K., Mitsuyama,S.,
Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordstiek,G.,
Hornischer,K., Barandt,P., Scharfe,M., Schoen,O., Desario,A.,
Reichelt,J., Kauer,G., Bloecker,H., Ramser,J., Beck,A., Klages,S.,
Hennig,S., Riesselmann,L., Dagand,E., Wehrmaeyer,S., Borzym,K.,


```
repeat_region 11567..11608
/Note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
repeat_region 11574..11618
/Note="MIR"
/rpt_family="SINE/MIR"
/rpt_type=DISPERSED
repeat_region 11625..11680
/Note="MIR"
/rpt_family="SINE/MIR"
/rpt_type=DISPERSED
repeat_region 11681..11978
/Note="AluJb"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
repeat_region 11979..12068
/Note="MIR"
/rpt_family="SINE/MIR"
/rpt_type=DISPERSED
repeat_region 13530..13835
/Note="AluSx"
/rpt_family="SINE/Alu"

Query Match 1.4%; Score 22; DB 9; Length 340000;
Best Local Similarity 100.0%; Pred. No. 0.8;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1177 ggatgtatatttctaggttcta 1198
|||||
Db 156788 GGATGTATATTCTAGGTCTA 156767

RESULT 9
HSPACTIXG/c
LOCUS HSPACTIXG 11873 bp DNA linear PAT 02-MAR-1995
DEFINITION H.sapiens factor IX gene.
ACCESSION A22491
VERSION A22491.1 GI:825659
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 11873)
AUTHORS Brownlee,G.G. and Choo,K.H.
TITLE Molecular cloning of the gene for human anti-haemophilic factor IX
JOURNAL Patent: EP 0107278-A 13 02-MAY-1984;
NATIONAL RESEARCH DEVELOPMENT CORPORATION
FEATURES
source
1..11873
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 3617 a 2081 c 2165 g 3180 t 830 others
ORIGIN

Query Match 1.3%; Score 21; DB 6; Length 11873;
Best Local Similarity 100.0%; Pred. No. 4.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 827 cagtggtcagtgcaatcataa 847
|||||
Db 1286 CAGTGGTCAGTGCATCATAA 1266

RESULT 10
HUMFIXG/c
LOCUS HUMFIXG 38059 bp DNA linear PRI 30-APR-1996
DEFINITION Human coagulation factor IX gene, complete cds.
ACCESSION K02402
VERSION K02402.1 GI:182612
KEYWORDS Alu repeat; Christmas factor; KpnI repetitive sequence;
```

```
antihemophilic factor B; factor IX; repeat region; simple
repetitive sequence.
Homo sapiens (clone: FIX-lambda-[6,36,53,61].) (tissue library:
T.Maniatis et al.) DNA; and Homo sapiens (clone: FIX-lambda-4243)
DNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 38059)
AUTHORS Yoshitake,S., Schach,B.G., Foster,D.C., Davie,E.W. and Kurachi,K.
TITLE Nucleotide sequence of the gene for human factor IX (antihemophilic
factor B)
JOURNAL Biochemistry 24 (14), 3736-3750 (1985)
MEDLINE 86000558
REFERENCE 2 (bases 23487 to 23556)
AUTHORS Rees,D.J., Rizza,C.R. and Brownlee,G.G.
TITLE Haemophilia B caused by a point mutation in a donor splice junction
of the human factor IX gene
JOURNAL Nature 316 (6029), 643-645 (1985)
MEDLINE 85296286
REFERENCE 3 (bases 23378 to 23387)
AUTHORS Graham,J.B., Lubahn,D.B., Lord,S.T., Kirshtein,J., Nilsson,I.M.,
Wallmark,A., Ljung,R., Frazier,L.D., Ware,J.L., Lin,S.W.,
Stafford,D.W. and Bosco,J.
TITLE The Malmo polymorphism of coagulation factor IX, an immunologic
polymorphism due to dimorphism of residue 148 that is in linkage
disequilibrium with two other F.IX polymorphisms
JOURNAL Am. J. Hum. Genet. 42 (4), 573-580 (1988)
MEDLINE 88161064
REFERENCE 4 (sites)
AUTHORS Hirosewa,S., Fahner,J.B., Sallier,J.-P., Wu,C.-T., Lovrien,E. and
Kurachi,K.
TITLE Structural and functional basis of the developmental regulation of
human factor IX gene: factor IX Leyden
JOURNAL Unpublished (1990)
COMMENT Sequence for [1] kindly submitted on floppy by K.Kurachi,
05-AUG-1985.
[1] notes a potential TATA box (2939-2942) and polyadenylation
signal (35701-35706); and notes two start codons (downstream of the
start codon annotated below) that may be alternative and/or
preferred starts for the factor IX prepropeptide. Several tracts
of simple repetitive sequence are present [1], including regions
with the potential for hairpin and/or Z-DNA formation. [1]
describes six long open reading frames in the intron and on the
complementary strand.
FEATURES
source
Location/Qualifiers
1..38059
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="Xq26.3-q27.1"
/clone="FIX-lambda-4243"
/clone="FIX-lambda-[6,36,53,61]."
/cell_line="49,XXXX"
/tissue_lib="fibroblast"
/tissue_lib="T.Maniatis et al."
repeat_region 1..2596
/rpt_family="KpnI"
prim_transcript 2966..35722
/gene="F9"
/note="FIX mRNA"
2966..35722
/gene="F9"
2966..3082
/gene="F9"
/note="G00-119-900"
/number=1
join(2995..3082,9291..9454,9643..9667,13357..13470,
20634..20762,23328..23530,33004..33118,33787..34334)
/gene="F9"
/note="precursor"
/codon_start=1
/product="factor IX"
/protein_id="AAB59620.1"
```

/db_xref="GI:182613"
/db_xref="GDB:G00-119-900"
/translation="MQRYNMIMAEPSGLITICLLGYLLSAECTVFLDHENANKILNRP
KRYNSCKLEEFVQGNLERECMEKCSFEAREVEFENTERTTTFWKQYVDGQCESNPC
LNGSCKDDINSYECWCPFGFEGKNCELDTCNIKGRCEQFCRNSADNKVVCSTEG
YRLAENQKSCPAVPFCGRVSVSQTSLTRAFAVFPDQVYNSTEAETILDNITOST
QSFNDFRVVGGEDAKPGQFPWQVVLNGKVDAFCCGGSIVNEKWLVTAAHCVETGVKIT
VVAGEHNIEFTEHTEQRNVIRIIPHHNVNAAINKYNHDIALLLEDEPLVLSYVTP
CIADKEYTNIFLFGSGYVSGWGRVFEHKGKRSALVQLYLRVPLVDRATCLRSTKFTIYN
NMFCAFHGGGRDSCQDGGPHVTEVEGTSFLTGTIISWGECAAMGKGYIYTKVSR
VNWIKETKLT"
join(2995. .3082,9291. .9340)
/gene="F9"
/note="G00-119-900"
3083. .9290
/gene="F9"
/note="G00-119-900"
/number=1
7298. .7593
/rpt_family="Alu"
8469. .8520
/gene="F9"
/note="t. .50 bp. .a in FIX-lambda-4243; ta in
FIX-lambda-[36,61]; G00-119-900"
/citation=[1]
/replace="ta"
9291. .9454
/gene="F9"
/note="G00-119-900"
/number=2
join(9341. .9454,9643. .9667,13357. .13470,20634. .20762,
23328. .23380)
/gene="F9"
/note="G00-119-900"
/product="factor IX light chain"
9455. .9642
/gene="F9"
/note="G00-119-900"
/number=2
9643. .9667
/gene="F9"
/note="G00-119-900"
/number=3
9668. .13356
/gene="F9"
/note="G00-119-900"
/number=3
10041
/gene="F9"
/note="g in one allele; c in another allele (loss of XmnI
recognition pattern); G00-119-900"
/replace="c"
13357. .13470
/gene="F9"
/note="G00-119-900"
/number=4
13471. .20633
/gene="F9"
/note="G00-119-900"
/number=4
14076. .14079
/gene="F9"
/note="tcga in one allele; nnnn in another allele (loss of
TaqI site); G00-119-900"
/replace="nnnn"
18165. .20265
/rpt_family="Kpni"
20416
/gene="F9"
/citation=[4]
/replace="a"
20634. .20762
/gene="F9"

/note="G00-119-900"
/number=5
20763. .23327
/gene="F9"
/note="G00-119-900"
/number=5
23328. .23530
/gene="F9"
/note="G00-119-900"
/number=6
23387
/gene="F9"
/note="g in [1]; a in [3] Ala->Thr; G00-119-900"
/replace="a"
23387
/gene="F9"
/note="g in one allele; a in another allele"
/replace="a"
join(23486. .23530,33004. .33118,33787. .34331)
/gene="F9"
/note="G00-119-900"
/product="factor IX heavy chain"
23531. .33003
/gene="F9"
/note="G00-119-900"
/number=6
23531
/gene="F9"
/note="g in [1]; t in [2] (haemophilia patient);
G00-119-900"
/replace="t"
24172. .24475
/rpt_family="Alu"
25863. .26091
/rpt_family="Alu"
31537. .31809
/rpt_family="Alu"
33004. .33118
/gene="F9"
/note="G00-119-900"
/number=7
33119. .33786
/gene="F9"
/note="G00-119-900"
/number=7
33787. .35722
/gene="F9"
/note="G00-119-900"
/number=8

Query Match 1.3%; Score 21; DB 9; Length 38059;
Best Local Similarity 100.0%; Pred. No. 3.8;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 827 cagtggtcagtgcaatcataa 847
|||||
Db 17480 CAGTGGTCAGTGCATCATAA 17460

RESULT 11
AL645665/c
LOCUS AL645665 153615 bp DNA linear HTG 01-FEB-2002
DEFINITION Homo sapiens chromosome X clone RP11-963P9, *** SEQUENCING IN
PROGRESS ***, 3 unordered pieces.
ACCESSION AL645665
VERSION AL645665.9 GI:18491387
KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (sites)
AUTHORS Heath,P.

TITLE Direct Submission
JOURNAL Submitted (31-JAN-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
COMMENT On Feb 4, 2002 this sequence version replaced gi:18476880.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: BA963P9
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 152173 bases at least Q40
Consensus quality: 152586 bases at least Q30
Consensus quality: 152764 bases at least Q20
Insert size: 153415; sum-of-contigs
Insert size: 144208; 38.0% error; agarose-fp
Quality coverage: 20.53x in Q20 bases; sum-of-contigs Quality coverage: 22.15x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.
* 1 28734: contig of 28734 bp in length
* 28735 28834: gap of 100 bp
* 28835 44671: contig of 15837 bp in length
* 44672 44771: gap of 100 bp
* 44772 153615: contig of 108844 bp in length.
FEATURES
source
1. .153615
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/clone="RP11-963P9"
/clone_lib="RPCI-11.4"
misc_feature
1. .28734
/note="assembly_fragment:06384
fragment_chain:1
clone_end:SP6
vector_side:left"
28835..44671
/note="assembly_fragment:02107
fragment_chain:1"
44772..153615
/note="assembly_fragment:00800
fragment_chain:1"
BASE COUNT 50474 a 29730 c 28392 g 44819 t 200 others
ORIGIN
Query Match 1.3%; Score 21; DB 2; Length 153615;
Best Local Similarity 100.0%; Pred. No. 3.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 827 cagtggtcagtgcgaatcataa 847
|||||||
Db 93977 CAGTGGTCAGTGCATCATAA 93957
RESULT 12
HS88D7/c
LOCUS HS88D7 158557 bp DNA linear PRI 23-NOV-1999
DEFINITION Human DNA sequence from clone 88D7 on chromosome Xq25-26.3 Contains F9 (coagulation factor IX (plasma thromboplastic component,

Christmas disease, haemophilia B)), db1 oncogene. EST, STS, GSS, complete sequence.
AL033403
AL033403.1 GI:3859054
HTG: Christmas factor; db1 oncogene; F9; factor IX; glycoprotein; oncogene; phosphoprotein; proto-oncogene.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 158557)
Bird,C.
Direct Submission
Submitted (09-NOV-1998) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence is the entire insert of clone 88D7.
This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.
This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/ChrX
88D7 is from the library RPC16 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see http://bacpac.med.buffalo.edu/ VECTOR: ppAC4.
FEATURES
Location/Qualifiers
1. .158557
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="q25-26.3"
/clone="RP6-88D7"
/clone_lib="RPCI-6"
2. .3009
/note="L1PA2 repeat: matches 3136.. 6144 of consensus"
3023.. 3684
/note="L1M4 repeat: matches 4639.. 5306 of consensus"
3693.. 3775
/note="L1MC/D repeat: matches 5517.. 5601 of consensus"
Join(<4083.. 4171,10329.. 10492,10681.. 10705,14395.. 14508, 21682.. 21810,24381.. 24583,34060.. 34174,34843.. 36777)
/gene="F9"
/note="match: DNA M21757 J00137 A13997 M33826 M11309 J00136 A22493 M23109 M35672 U51135 M26247 M26233 M26234 M26235 M26236 M26237 Z50155; match: STS G28615"
/evidence=not_experimental
4083.. 36777
/gene="F9"
CDS
Join(4099.. 4171,10329.. 10492,10681.. 10705,14395.. 14508, 21682.. 21810,24381.. 24583,34060.. 34174,34843.. 35390)
/gene="F9"
/note="match: protein P00740 P19540 P16294 P00741 P16293 P16296 P16295 P16292 P16291 P25155"
/codon_start=1
/evidence=not_experimental
/product="F9 (coagulation factor IX (plasma thromboplastic component, Christmas disease, haemophilia B))"
/protein_id="CAA21954.1"
/db_xref="GI:3859055"
/db_xref="SPTREMBL:O14316"
/translation="MNAESPLGITICLLGYLLSAECTVFLDHENANKILNRPKRYNS GKLEEFVQGNLERECMEKCSFEAREVFENTERTTEFWKQYVDGQCESNPCLNGGS

CKDDINSYECWCPGFEKNCNCELDVTCNIKNGRCEQFCCKNSADNKVVCSCTEGYRLAE
NQKSCPAVPFCGRVSVSQTSKLTRAETVFPDQVDVYNSTEAETILDNITQSTQSPND
FTRVGGEDAKPGQFPQWVVLNGKVDAPFCGGSIVNEKWIYVTAACHCVETGKRTVVAGE
HNIETEHTECKRNVIRIIPHNNYNAINKYNHHDIALLEDEPLVLSYVTPICIADK
EYTNIFLKFGSGYVSGWRVFKGRSALVQLYLRVPLVDRATCLRSTKFTIYNMFCA
GFHEGGRDSCQGDGSGPHVTEVEGTSFLTGIISWGECAMKGYCIYTKVSRVYNWIK
EKTGKLT

repeat_region 5799. .6479 /note="L2 repeat: matches 1963. .2708 of consensus"
repeat_region 6562. .5807 /note="MIR repeat: matches 6. .262 of consensus"
repeat_region 6829. .6925 /note="L2 repeat: matches 1641. .1739 of consensus"
repeat_region 6965. .7015 /note="L2 repeat: matches 2652. .2700 of consensus"
repeat_region 7150. .7512 /note="L2 repeat: matches 1244. .1617 of consensus"
repeat_region 7565. .7764 /note="MIR repeat: matches 14. .198 of consensus"
repeat_region 7771. .8111 /note="MER2 repeat: matches 1. .345 of consensus"
repeat_region 8388. .8685 /note="AluSx repeat: matches 1. .292 of consensus"
repeat_region 9091. .9431 /note="MER7A repeat: matches 2. .346 of consensus"
repeat_region 9468. .9609 /note="71 copies 2 mer at 78% conserved"
repeat_region 10959. .11604 /note="L2 repeat: matches 2123. .2746 of consensus"
repeat_region 12270. .12305 /note="18 copies 2 mer aa 81% conserved"
repeat_region 12516. .12693 /note="MIR repeat: matches 9. .190 of consensus"
repeat_region 14736. .14864 /note="MIR repeat: matches 52. .193 of consensus"
repeat_region 17136. .17189 /note="L2 repeat: matches 2645. .2706 of consensus"
repeat_region 17940. .18249 /note="LTR16C repeat: matches 58. .383 of consensus"
repeat_region 18718. .19054 /note="L1MEC repeat: matches 2414. .2414 of consensus"
repeat_region 19251. .19679 /note="L1PA14 repeat: matches 5718. .6149 of consensus"
repeat_region 19857. .20579 /note="L1M4 repeat: matches 4181. .4940 of consensus"
repeat_region 20625. .20789 /note="L1M4 repeat: matches 5079. .5249 of consensus"
repeat_region 20788. .21386 /note="L1ME3 repeat: matches 5517. .6155 of consensus"
repeat_region 22684. .23154 /note="MER9 repeat: matches 1. .511 of consensus"
repeat_region 25061. .25226 /note="MER5A repeat: matches 16. .186 of consensus"
repeat_region 25227. .25532 /note="AluSx repeat: matches 1. .302 of consensus"
repeat_region 25840. .26048 /note="MER3 repeat: matches 2. .208 of consensus"
repeat_region 26919. .27152 /note="AluY repeat: matches 69. .302 of consensus"
repeat_region 31078. .32008 /note="MER45C repeat: matches 1. .948 of consensus"
repeat_region 32082. .32243 /note="MER5A repeat: matches 10. .182 of consensus"
repeat_region 32254. .32466 /note="MIR repeat: matches 14. .252 of consensus"
repeat_region 32592. .32859 /note="AluSc repeat: matches 1. .291 of consensus"
repeat_region 32937. .32978 /note="L2 repeat: matches 2646. .2691 of consensus"
repeat_region 37001. .37314 /note="AluSx repeat: matches 1. .302 of consensus"
repeat_region 37791. .37847 /note="L2 repeat: matches 2648. .2706 of consensus"

repeat_region 39138. .39181 /note="L2 repeat: matches 2664. .2707 of consensus"
repeat_region 39985. .40256 /note="AluSx repeat: matches 1. .273 of consensus"
repeat_region 40837. .41091 /note="AluJb repeat: matches 41. .295 of consensus"
repeat_region 41847. .41867 /note="MER57B repeat: matches 381. .403 of consensus"
repeat_region 41868. .42320 /note="L1MA2 repeat: matches 5843. .6308 of consensus"
repeat_region 42321. .42472 /note="MER57B repeat: matches 250. .381 of consensus"
prim_transcript 42331. .>43518 /note="match: multiple ESTs; match: 5' EST H94135 clone
242834; Paired with EST H94060 matching this clone; match:
3' EST H94060 clone 242834; Paired with EST H94135
matching this clone"
repeat_region 42343. .42499 /note="MER93 repeat: matches 216. .371 of consensus"
repeat_region 42574. .42644 /note="MER93 repeat: matches 18. .81 of consensus"
repeat_region 42651. .42941 /note="AluSx repeat: matches 6. .296 of consensus"
repeat_region 42965. .43128 /note="MER63 repeat: matches 1. .789 of consensus"
repeat_region 43519. .43554 /note="18 copies 2 mer aa 86% conserved"
repeat_region 43944. .44135 /note="L1ME3A repeat: matches 5673. .5859 of consensus"
repeat_region 44701. .44995 /note="AluJb repeat: matches 1. .291 of consensus"
repeat_region 45878. .46458 /note="L1PA7 repeat: matches 5563. .6145 of consensus"
repeat_region 46626. .47041 /note="MSTA repeat: matches 1. .426 of consensus"
repeat_region 50779. .50810 /note="16 copies 2 mer tc 84% conserved"
repeat_region 51052. .51080 /note="L1M4 repeat: matches 5101. .5127 of consensus"
repeat_region 51081. .51286 /note="L1PB1 repeat: matches 5936. .6151 of consensus"
repeat_region 51287. .52485 /note="L1M4 repeat: matches 3846. .5101 of consensus"
repeat_region 52496. .52839 /note="L1M4 repeat: matches 2644. .3028 of consensus"
repeat_region 53386. .53417 /note="16 copies 2 mer gt 91% conserved"
repeat_region 54396. .54454 /note="MIR repeat: matches 201. .257 of consensus"
repeat_region 54403. .54462 /note="L2 repeat: matches 2654. .2748 of consensus"

Query Match 1.3%; Score 21; DB 9; Length 158557;
Best Local Similarity 100.0%; Pred.No. 3.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 827 cagtgtcagtgcaatcataa 847
|||||
Db 18525 CAGTGTCTCAGTGCATCATAA 18505

RESULT 13
AF052582
LOCUS AF052582 2154 bp DNA linear PLN 12-APR-1999
DEFINITION Fusarium oxysporum f. sp. lycopersici 42-87 family F xylanase
(xyl3) gene, complete cds.
ACCESSION AF052582
VERSION AF052582.1 GI:2981134
KEYWORDS
SOURCE Fusarium oxysporum f. sp. lycopersici.
ORGANISM Fusarium oxysporum f. sp. lycopersici
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreales; Mitosporic Hypocreales; Fusarium; Fusarium oxysporum

complex.
1 (bases 1 to 2154)
Ruiz-Roldan,M.C., Di Pietro,A., Huertas-Gonzalez,M.D. and
Roncero,M.I.
Two xylanase genes of the vascular wilt pathogen Fusarium oxysporum
are differentially expressed during infection of tomato plants
Mol. Gen. Genet. 261 (3), 530-536 (1999)
99254813
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
2 (bases 1 to 2154)
Ruiz-Roldan,M.C., Huertas-Gonzalez,M.D., DiPietro,A. and
Roncero,M.I.G.
Direct Submission
Submitted (08-MAR-1998) Dept. Genetica, Facultad de Ciencias,
Universidad de Cordoba, Avda San Alberto Magno, Cordoba 14071,
Spain
FEATURES
source Location/Qualifiers
1. .2154
/organism="Fusarium oxysporum f. sp. lycopersici"
/strain="42-87"
/db_xref="taxon:59765"
409. .1670
/gene="XYL3"
/join(409. .481,537. .863,916. .1670)
/gene="XYL3"
/codon_start=1
/product="family F xylanase"
/protein_id="AAC06239.1"
/db_xref="GI:2981135"
/translation="MHTFVLLALAPVSALAAQPIWQCGGNGWTGATTCASGLKCEK
INDWYQVPGSGGSEPPQSSTQGGTPOPTGGNSGGTGLDAKFKAGKQYFGTEIDH
YHLNNPLINIVKQFGQVTCENSMKWDAIEPSRNSFTFSNADKVVDPATONGKLIRG
HTLLWHSOLPQWQVONINDRSTLTAVIENHVKTMTVTRYKGKILQWDVVNEIFAEDGNLR
DSVFSVLGEDFVGIAFRAARAADPAAKLYINDYNDKSDYAKLTRGMVAHVNKWIAA
GIPIDGIGSQGHLAAPSGWNPASGVPAALRALAASDAKEIAITELDISGASANDYLTV
MNACLAVPKCVGITVWGVSDKDSWRPGENPLLYDSNYQPKAAFNALVNAL"
BASE COUNT 509 a 597 c 502 g 546 t
ORIGIN
Query Match 1.3%; Score 20; DB 8; Length 2154;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1204 gcttcaatgccttggtcaa 1223
|||||
Db 1641 GCTTCAATGCCTTGGTCAA 1660
RESULT 14
RNPFKL
LOCUS RNPFKL 2770 bp mRNA linear ROD 16-DEC-1991
DEFINITION Rat PFK-L mRNA for liver phosphofructokinase.
ACCESSION X58865
VERSION X58865.1 GI:56886
KEYWORDS phosphofructokinase.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 2770)
AUTHORS Hotta,K.
TITLE Direct Submission
JOURNAL Submitted (03-APR-1991) K. Hotta, Second Dept of Internal Medicine,
Osaka University Medical School, 1-1-50 Fukushima, Fukushima-ku,
Osaka 553, JAPAN
REFERENCE 2 (bases 1 to 2770)
AUTHORS Hotta,K., Nakajima,H., Yamasaki,T., Hamaguchi,T., Kuwajima,M.,
Noguchi,T., Tanaka,T., Kono,N. and Tarui,S.
TITLE Rat-liver-type phosphofructokinase mRNA. Structure, tissue
distribution and regulation
JOURNAL Eur. J. Biochem. 202 (2), 293-298 (1991)
MEDLINE 92104147

FEATURES
source Location/Qualifiers
1. .2770
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="lambda-PFKL1-PFKL8"
/tissue_type="liver"
/clone_lib="lambda-gt10"
1. .2740
/gene="PFK-L"
/evidence=experimental
1. .2740
/gene="PFK-L"
62. .70
/gene="PFK-L"
67. .2409
/gene="PFK-L"
/EC_number="2.7.1.11"
/codon_start=1
/product="6-phosphofructokinase"
/protein_id="CAA41674.1"
/db_xref="GI:56887"
/db_xref="SWISS-PROT:P30835"
/translation="MATVDLEKLRMSGAGKAIGVLTSGGDAQGMNAAVRAVTRMGIVV
GAKFLIYEGVEGLVEGENIKPANWLSVSNIIQLGGTIGSARCKAFTTREGRLAAA
YNLLQGITNLCVIGGDSLTGANIFRNEWGSLLEELVKEGKISESTAQNYAHLISAG
LVGSINDFCGDMTIGTDSALHRIEMVIDAITTAQSHQRTFVLEVMGRHCGYIALV
SALASGADWLFIEAPPEDGCWENFNCERLGETRSRGSRLNIIIAEGAIDRHGKPISS
SYVKDLVVQLRGFDTRVTVLGHVQGGTSPAFORVLSKKNMEAVMALLEATPDPAC
VVSLSGNQSVRLPLMECVQVTKDQKAMDEKRFDEAIQLRGRSFENNWKIYKLLAHOK
VSKERNSFLAIIINVGAPAGMNAAVRSARVTGISSEHTYVVVDGFEGLARGOVQEV
GWHDVAGWLGSGSMLGTRTKLPKPLEAIVENLRTYNIHALLVIGGFEAYEGLQLV
EARGRYELCLVMCVIPATISNNVPGTDFSLGSDYAVNAAMESCDRIKQASAGTKRRV
FIVETMGYCGYLATVTGIAVGADAAVVFEDPENIHDLKANVEHMTKMKTDIQRGLV
LRNEKCHEHYTTEFLYNIYSSSEGRGVDFCRTNVLHLOQGGAPTFFDRNYGTLGVKA
MLWMSEKLRDYYRKGRVFEANPDSACVIGLRKKVAVFSSVTELKKTDFEHRMPREQW
WLNRLRLMKMLAHYRISMADYVSGELEHVTRRTLSIDKGF"
2722. .2727
/gene="PFK-L"
2740
/gene="PFK-L"
BASE COUNT 609 a 762 c 864 g 535 t
ORIGIN
polyA_signal
polyA_site
Query Match 1.3%; Score 20; DB 10; Length 2770;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 505 gtcttcctcatctacgagg 524
|||||
Db 208 GTCTTCTCATCTACGAGG 227
RESULT 15
BC007536
LOCUS BC007536 2879 bp mRNA linear PRI 12-JUL-2001
DEFINITION Homo sapiens, phosphofructokinase, liver, clone MGC:15386
IMAGE:3347301, mRNA, complete cds.
ACCESSION BC007536
VERSION BC007536.1 GI:14043100
KEYWORDS MGC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 2879)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (10-MAY-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

REMARK
COMMENT

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTP
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@bcgsc.bc.ca
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL Plate: 22 Row: 1 Column: 1
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis, Genomescan gene prediction, similarity but not identity
to protein.

FEATURES
source

Location/Qualifiers
1. .2879
/organism="Homo sapiens"
/db_xref="LocusID:5211"
/db_xref="taxon:9606"
/clone="MGC:15386 IMAGE:3347301"
/tissue_type="Kidney, renal cell adenocarcinoma"
/clone_lib="NIH_MGC_14"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
16. .2358

CDS

/codon_start=1
/product="phosphofructokinase, liver"
/protein_id="AAH07536.1"
/db_xref="GI:14043101"
/translation="MAAVDLEKLRASGAGKAIGVLTSGGDAQGMNAAVRAVTRMGIYV
GAKVELIYEGYGLVEGGENIKQANWLSVSNIIQLGGTIIGSARCKAFTTREGRRAAA
YNLVQHGITNLCVIGGSLTGANIFRSENGSLLEELVAEGKISSETTARTYSHLNIAG
LVGSDNDFCGTDMTIGTDSALHRIMEVIDAITTAQSHQRTFVLEVMGRHCGYLALV
SALASGADWLFIFEAPPDGEWFMFCERLGETRSRSLNIIIAEGAIDRNGKPISS
SYKDLVVQRLGFDTRTVVLGHVQGGTPSAFDRILSSKMGMEAVMALLEATPDTTAC
VVTLSGNQSVRLPLMECVQMTKEVQKAMDDKREDEATQLRGGSFENNWNIIYKLLAHQK
PPKESNFSLAILNVGAPAGMNAAVRSVVRGISHGHTVYVVDGPEGLAKGVQEV
GWHVAGWLGKGGMLGTRTLPKQLESIVENIRIYGIHALLVVGGEAYEGVLQLV
EARGRYEELCIVMCVIPATISNNVPGTDFSLGSDTAVNAAMESCDRIKQASAGTKRRV
FIVETMGYCGYLATVTGIAVGADAAYVFEDPFNHDLKVNVEHMTKMKTDIQRGLV
LRNEKCHDYTTTEFLYNLYSSEKGVDFDRTNVLHLOQGGAPTPEEDNYGTLGVKA
MLWLSKLRVYRKGRVFANAPDSACVIGLKKKAVAFSPVTELKKTDFEHRMPREQW
WLSLRMLKMLAQYRISMAAYVSGELEHVTTRRLSMDKGF"

BASE COUNT
ORIGIN

554 a 875 c 932 g 518 t

Query Match 1.3%; Score 20; DB 9; Length 2879;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 505 gctctctcatctacgaggg 524
|||||
Db 157 GTCTCTCATCTACGAGG 176

Search completed: April 27, 2002, 05:44:33
Job time: 17910 sec


```
Db 146 YRYFAGWADKWHGKTIIPMDGQHCFTRHEPVGCGQIIPWNFPLVMQGWKLAPALATGNT 205
      :||:||||| ||| | | : ||:|||| | || : ||:||||| | :
QY 178 FIFKASEKSPGLVGLAPFAEAGFPVQVQLTGART-GEALASHMDIAKISFTRSVG 236
      : ||: ||| | | | | ||| ||| : | | | ||| : ||: ||| |
Db 206 VMKVAEQTPLSALYLASLIKEAGFPVGVNIITGYGTAGAAIAOHMDVDKVAFTGSTE 265
QY 237 GGRAVKQATLKSNNKRVTLLEGEK-PTIVFNEAPLE---RQSGESAKDFSFGQIWWPPS 292
      | : ||: ||| ||| ||| | ||| : | | : | | : | |
Db 266 VGHLOKAAGDSNLKRVTLLEGGKSPSIVLADADMEHAVEQCHEAL--FFNMGQCCAGS 323
QY 293 CLLVQWG-----NLAERFHGVRHGSFGGCGQWLGNLGNPLEPKRTHGPFVDKSOYDRVLG 345
      : | : ||: ||| ||| ||| | ||| : | | | | | ||| |||
Db 324 RTFVEESIYNEFLERTVEK-----AKQRKVG-NPFELDTQGGPQVDKEQFERVLG 372
QY 346 NIDVG-KDTAQLLTGVRKDGKFAIEPTIFVNPKPGSKIWFEEIFGPVLSIKTFKTEEE 404
      | : | : ||| | | | ||| ||| : | | ||| ||| : || ||
Db 373 YIQLGQKEGAKLLCGGERGEGFFIKPTVFGGVQDDMRIAKEEIFGPVQPLFKFKIEE 432
QY 405 ALETANDTTYGLASVIYTKSLNRGLRVSSALETGGVSINFPFIPETQTFFGGMKQSGSGR 464
      : | ||: ||| ||| : ||: ||: | | | | | ||| ||| |||
Db 433 VVERANNTRYGLAAAVFTRDLQKAMYFTQALQAGTVWYNTYNTVCHTPTFGGFKESNGR 492
QY 465 ELGEEGLKAYLEPKTINIHV 484
      ||||:|||| | ||: | |
Db 493 ELGEDGLKAYTEVKTIVK 512
```

```
RESULT 2
Q9C1Q6 PRELIMINARY: PRT: 497 AA.
AC Q9C1Q6;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE ALDEHYDE DEHYDROGENASE ALDH.
GN ALDA.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=5072;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21269214; PubMed=11102439;
RA Flippin M., Mathieu M., Cirpus I., Panozzo C., Felenbok B.;
RT "Regulation of the aldehyde dehydrogenase gene (aldA) and its role in
RT the control of the co-inducer level necessary for induction of the
RT ethanol utilization pathway in Aspergillus nidulans.";
RL J. Biol. Chem. 276:6950-6958(2001).
DR EMBL; AF260123; AAK18072.1; -.
DR HSSP; P05091; 1CW3.
DR InterPro; IPR002086; Aldehyde_dehydr.
DR Pfam; PF00171; aldedh; 1.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
FT VARIANT 191 191 E -> Q.
FT VARIANT 208 208 G -> P.
FT VARIANT 247 247 R -> P.
FT VARIANT 409 409 E -> V.
SQ SEQUENCE 497 AA; 54138 MW; 9C6CFC62731A25BD CRC64;
```

```
Query Match 34.9%; Score 889; DB 3; Length 497;
Best Local Similarity 43.1%; Pred. No. 8.3e-61;
Matches 207; Conservative 78; Mismatches 173; Indels 22; Gaps 11;

QY 12 LFINNEFVSSKGSERLTLPNPDESTVATDVHVANAADVDSAVAASVQAVKGPWKFTG 71
      ||||| | : ||: | | | | | ||| ||| : | : ||| : |
Db 21 LFINNEFVKGVEKTFQVINPSNEK-VITSVHEATEKDVDVAVAAA-RAAFEGPWRQVTP 78
      : ||: ||| ||| : ||: ||| ||| ||| ||| : ||: ||| : |
QY 72 AQRACMLKFADLAERNAEKLARLESPTGRPVSMITHFDIPNMVSVFRRYAGWADKIAG 131
      : ||: ||| ||| : ||: ||| ||| : | : | : ||| ||| |
Db 79 SERGILINKLADLMERDITLAAIESLDNGKAFM-AKVDLANSIGCLRRYAGWADKIING 137
```

```
QY 132 KTFPEDNGKPNW-RYEPMGVCAGIASWNATFLYVGWKIAPALAAAGCSFIFKASEKSPGLV 190
      : | : ||: ||| | | | | ||| ||| : | : ||| : |
Db 138 QTIDTNPETLTYYTRHEPVGCGQIIPWNFPLMWSWKIGPAVAAGNTVVLKTAEQTPLSA 197
QY 191 LGLAPLFAEAGFPVQVQLTG-ARVTGEALASHMDIAKISFTRSVGGGRAVKQATLKS 249
      | | | ||| ||| : ||: ||| ||| ||| ||| : ||: ||| |
Db 198 LYAAKLIKEAGFPAGVINVISGFGRTAGAAISSHMDIDKVAFTGTLVGRITLQAAAKSN 257
QY 250 MKRVTLELGEK-PTIVFNEAPLERQ-SGESAKDFSFGQIWWPPSCLLVQWGNLAEKF-- 305
      : ||: ||| ||| | | ||| ||| : | : | | | : |||
Db 258 LKKVTLELGGKSPNIVFDDADIDNAISWANFAGIFFNHGQCCAGSRILVQEG-IYDKFVA 316
QY 306 --HGVRHGSFGGCGQWLGNLGNPLEPKRTHGPFVDKSOYDRVLGNIDVGKDT-AQLLTGVG 361
      : | : ||: ||| ||| ||| ||| ||| ||| ||| : ||: |||
Db 317 RPKERAQKNKVG-----NPFQDFTFGPQVSQLQFDRIMEYINHGKKAGATVATGGD 368
QY 362 RKDGKGAIEPTIFVNPKPGSKIWFEEIFGPVLSIKTFKTEEEAIEIANDTTYGLASVIY 421
      | ||: ||| ||| : ||| ||| ||| ||| ||| ||| ||| : ||: |||
Db 369 RHGNEYFIQPTVFTDVTSDMKIAQEEIFGPVVTIQKFDEAEAIKIGNSTDYGLAAAVH 428
QY 422 TKSINRGLRVSSALETGGVSINFPFIPETQTFFGGMKQSGSRELGEGLKAYLEPKTIN 481
      ||: ||| ||| ||| ||| : ||: ||| ||| ||| ||| ||| : ||: |||
Db 429 TKNVNTAIRVSNALKAGTVWINNINMISYQAPFGGFKQSGLRELGSYALENYTOIKTVH 488

RESULT 3
Q9C1Q5 PRELIMINARY: PRT: 497 AA.
AC Q9C1Q5;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE ALDEHYDE DEHYDROGENASE ALDH15.
GN ALDA.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=5072;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21269214; PubMed=11102439;
RA Flippin M., Mathieu M., Cirpus I., Panozzo C., Felenbok B.;
RT "Regulation of the aldehyde dehydrogenase gene (aldA) and its role in
RT the control of the co-inducer level necessary for induction of the
RT ethanol utilization pathway in Aspergillus nidulans.";
RL J. Biol. Chem. 276:6950-6958(2001).
DR EMBL; AF260124; AAK18073.1; -.
DR HSSP; P05091; 1CW3.
DR InterPro; IPR002086; Aldehyde_dehydr.
DR Pfam; PF00171; aldedh; 1.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
FT VARIANT 286 286 V -> A.
SQ SEQUENCE 497 AA; 54166 MW; 65849D1DCA94A5F5 CRC64;
```

```
Query Match 34.8%; Score 888; DB 3; Length 497;
Best Local Similarity 43.1%; Pred. No. 9.9e-61;
Matches 207; Conservative 78; Mismatches 173; Indels 22; Gaps 11;

QY 12 LFINNEFVSSKGSERLTLPNPDESTVATDVHVANAADVDSAVAASVQAVKGPWKFTG 71
      ||||| | : ||: | | | | | ||| ||| : | : ||| : |
Db 21 LFINNEFVKGVEKTFQVINPSNEK-VITSVHEATEKDVDVAVAAA-RAAFEGPWRQVTP 78
      : ||: ||| ||| : ||: ||| ||| ||| ||| : ||: ||| : |
QY 72 AQRACMLKFADLAERNAEKLARLESPTGRPVSMITHFDIPNMVSVFRRYAGWADKIAG 131
      : ||: ||| ||| : ||: ||| ||| : | : | : ||| ||| |
Db 79 SERGILINKLADLMERDITLAAIESLDNGKAFM-AKVDLANSIGCLRRYAGWADKIING 137
QY 132 KTFPEDNGKPNW-RYEPMGVCAGIASWNATFLYVGWKIAPALAAAGCSFIFKASEKSPGLV 190
      : | : ||: ||| | | | | ||| ||| : | : ||| : |
Db 138 QTIDTNPETLTYYTRHEPVGCGQIIPWNFPLMWSWKIGPAVAAGNTVVLKTAEQTPLSA 197
```


QY 121 YVAGWADKIAGKTFPEDNGKPNW-RYEPMGVCAGIASWNATFLYVGWKIAPALAAAGCSFI 179
Db 150 YFAGWADKWHGKTIPTMDGEHFCFTRHEPVGCGOIIPWNFPLVMGQWKLAPALATGNITVV 209
QY 180 FKASKSPLGVGLAPLFAEAGFPVGVVQVLTGARVT-GEALASHMDIAKISFTRSVGGG 238
Db 210 MKVAEQTPLSALYLASLIKEAGFPVGVVNIITGYGPTAGAAIAQHMDVDKVAFTGSTEVG 269
QY 239 RAVKOATLKSNMKRVTLLEGEK-PTIVFNEAPLER--QSGESAKDFSFKFGQIWVPPSCL 294
Db 270 HLIQKAAGESNLKRVTLLEGGKSPSIVLADADMEHAVDQCHEAL--FFNMGQCCCAGSRT 327
QY 295 LVWVG-----NLAERFHGVRHGSFGGCGORWLGONPLEPKRTHGPFVDKSOYDRVLGNI 347
Db 328 FVEBSIYREFLERTVEK-----AKQKVG-NPFELDTQOGPQVDKEQFERILGYI 376
QY 348 DVG-KDTAQLLTGVGRKDGKGAIEPTIFVNPKPGSKIWFEEIFGPVLSIKTFKTEEBAI 406
Db 377 RLGOKEGAKLLCGGERLGERGFFIKPTVFGDVQDGMRIAKEEIFGPVQPLFKFKKIBEVI 436
QY 407 ETANDTTYGLASVIYTKSLNRGLRVSSALETGGVSINFFPIPETQTPFGGMKQSGSREL 466
Db 437 QRANTRYGLAAAVFTRDLDKAIYFTQALQAGTVVWNTYNIVTCHTPFGGFKESGNGREL 496
QY 467 GEGLKAYLEPKTINIHV 484
Db 497 GEDGLRAYTEKVTITKV 514
RESULT 6
O35945
ID O35945 PRELIMINARY; PRT; 501 AA.
AC O35945;
DT 01-JAN-1998 (TremBLrel. 05, Created)
DT 01-JAN-1998 (TremBLrel. 05, Last sequence update)
DT 01-OCT-2001 (TremBLrel. 18, Last annotation update)
DE ALDEHYDE DEHYDROGENASE AHD-2-LIKE.
GN ALDH1A4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL; TISSUE=LIVER;
RA Hsu L.C., Chang W.C., Yoshida A.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U96401; AAB64411.1; -.
DR HSSP; P51977; 1BX5.
DR MGD; MGI:1347050; Aldh1a4.
DR InterPro; IPR002086; Aldehyde_dehydr.
DR Pfam; PF00171; aldedh; 1.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
SQ SEQUENCE 501 AA; 54587 MW; 42DE97962799237B CRC64;

Query Match 34.5%; Score 878.5; DB 11; Length 501;
Best Local Similarity 41.9%; Pred. No. 5.5e-60;
Matches 204; Conservative 89; Mismatches 173; Indels 21; Gaps 11;

QY 10 SELFIFNNEFVSSKGSERLTLTNPDWDESTVATDVHVANAADVDSAVAASVQAVKKG-PWKK 68
Db 21 TKIFINNEHDSVSSKKFPVLPNATEEVIC-HVEEGDKADVDAKAVKAAQAFQIGSPWRT 79
QY 69 FTGAQRAACMLKPADLAENAKELARLESPTGRPVSMITHFDIPNMVSVFRYVAGWADK 128
Db 80 MDASERGLLNKLADLMERDRLLATMESMNAGKVFAHAHYLIDVEISIKALQYFAGWADK 139
QY 129 IAGKTFPEDNGKPNW-RYEPMGVCAGIASWNATFLYVGWKIAPALAAAGCSFIFKASEKSP 187
Db 140 IHGQTIIPSDGNIIFTYTRREPIGVCGIIPWNGPLIIFTWKILGPALSCGNTVVVPAEQTP 199

QY 188 LGVLGLAPLFAEAGFPVGVVQVLTGARVT-GEALASHMDIAKISFTRSVGGRAVKQATL 246
Db 200 LTALHMASLIKEAGFPVGVVNIIVPGYGTAGGAISSHMDIDKVSFTGSTGVKLIKEAAG 259
QY 247 KSNMKRVTLLEGEK-PTIVFNEAPLERQSGESAKDFS-----KFGQIWVPPSCLLVQWG 299
Db 260 KSNLKRVTLELGGKSPCIVFADADL----DSAVEFAHOGVFFHQGOICVAASRLFEVE-E 313
QY 300 NLAERFHGVRHGSFGGCGORWLGONPLEPKRTHGPFVDKSOYDRVLGNIDVG-KDTAQLLT 358
Db 314 SIYDEF--VRR-SVERAKKYILGNPLNSGINOGPOIDKEQHNKILGLIESGKKEGAKLEC 370
QY 359 GVGRKDGKGAIEPTIFVNPKPGSKIWFEEIFGPVLSIKTFKTEEEAIEIANDTTYGLAS 418
Db 371 GGRWGNKGFFVQPTVFSNVTDGMRIAKEEIFGPVQIMFKSMDDVIKRANNTTYGLAA 430
QY 419 VIYTKSLNRGLRVSSALETGGVSINFFPIPETQTPFGGMKQSGSRELGEGLKAYLEPK 478
Db 431 GVFTKDLDKAITVSSALQAGMVWVNCYLAVPVQCPFGGFKMSGNGRELGHGLYEYTELK 490
QY 479 TINIHN 485
Db 491 TVAMQIS 497
RESULT 7
Q9P303
ID Q9P303 PRELIMINARY; PRT; 497 AA.
AC Q9P303;
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE ALDEHYDE DEHYDROGENASE.
GN ALDH.
OS Cladosporium fulvum (Fulvia fulva).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina;
OC Dothideomycetes et Chaetothyriomycetes incertae sedis;
OC Mycosphaerellaceae; mitosporic Mycosphaerellaceae; Cladosporium.
OX NCBI_TaxID=5499;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CF4;
RX MEDLINE=98052127; PubMed=9390425;
RA Coleman M., Henricot B., Arnau J., Oliver R.P.;
RT "Starvation-induced genes of the tomato pathogen Cladosporium fulvum are also induced during growth in planta.";
RL Mol. Plant Microbe Interact. 10:1106-1109(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CF4;
RA Segers G.C., Bradshaw N., Archer D., Blissett K., Oliver R.P.;
RT "Alcohol oxidase is a novel pathogenicity factor for Cladosporium fulvum but aldehyde dehydrogenase is dispensable.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF275347; AAF82789.1; -.
DR HSSP; P05091; 1CW3.
DR InterPro; IPR002086; Aldehyde_dehydr.
DR InterPro; IPR001128; Cyt_p450.
DR Pfam; PF00171; aldedh; 1.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
SQ SEQUENCE 497 AA; 53849 MW; 1B1EEB10FB5599 CRC64;

Query Match 34.2%; Score 872; DB 3; Length 497;
Best Local Similarity 42.1%; Pred. No. 1.7e-59;
Matches 207; Conservative 74; Mismatches 181; Indels 30; Gaps 12;

QY 7 EYKSELFIFNNEFVSSKGSERLTLTNPDWDESTVATDVHVANAADVDSAVAASVQAVKKG 66
Db 16 EQPTGLFINNEWKVPVSGKFFDVFINTDESVC-QVAEADKDVDTAVAAARKAF-EGPW 73

QY	67	KKFTGAQRAACMLKFADLAEKNAEKIARLES	PTGRPVSMITHFDIPNMVSFVRYAGWA	126
Db	74	RQETPENRGKLLVLAELFEKNLDFASVKALONGKAFNM-AKVDIGMCAGCLRYYGGWA	132	
QY	127	DKIAGKTFPEDNGKPNWRY--EPMGVCA	GIASWNAFLYVGWKIAPALAAAGCSFIFKAS	183
Db	133	DKIEGKVV-DTTPDTFNYIKKEPIGVCGQIIPWNPFLMWAWKIGPAVATGNIVVIKTA	190	
QY	184	EKSPLGVGLGLAPLFAEAGFP	PGVYQFLTG-ARVTGEALASHMDIAKISFTRSVGGGRAVK	242
Db	191	EQTPLGAYAAANLIKEAGFP	PGVNVVISGFGKVAGAAIASHMDIDKVAFTGSTVTGTIL	250
QY	243	QATLKSNNKRVTL	ELGEK-PTIVFNEAPLERQ-SGESAKDFSKFGQIWVPPSCLLVQWGN	306
Db	251	KAAAGSNLKKVTL	ELGGKSPNIVFDDADIDNAISWVNEGIFFNHGQRCACGSRIVVQEG-	309
QY	301	LAKEFHGVRHGSFGGCQRWLQ-----NPLEPKRTHGPFVDKSOYDRVLGNIDVGDKT	353	
Db	310	IYDKF-----LQRFKERAQKNVVGDPFAADTFQGPQVSQVQFDRIMGYIQAGRDA	359	
QY	354	-AQLLTGVRKGDKGFAIEPTIVNPKPGSKINWEEIFGPVLSIKTEKTEEEAIEIANDT	412	
Db	360	GATVEIGNRKGDKGFIPTIFANVTEDMKIMQEEIFGPVCSISKFKTKEDAIKVGNA	419	
QY	413	TYGLASVITYTKSLNRGLRVSSALETGGVSINPFIPTOTPFPGMKQSGSGREIGEGLK	472	
Db	420	PYGLAAAVHTRNLNTAIEVSNALRAGTVVWVNTYNTLHHOLPFGGFKESGIGREJGEDALA	479	
QY	473	AYLEPKTINIHV	484	
Db	480	NYTOTKTQTSIRL	491	

```

RESULT      8
Q9TXM0
ID      Q9TXM0      PRELIMINARY;      PRT;      537 AA.
AC      Q9TXM0;
DT      01-MAY-2000 (TrEMBLrel. 13, Created)
DT      01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT      01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE      HYPOTHETICAL 58.6 KDA PROTEIN.
GN      K04F1.15.
OS      Caenorhabditis elegans.
OC      Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC      Rhabditidae; Peloderinae; Caenorhabditis.
OX      NCBI_TaxID=6239;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=BRISTOL N2;
RX      MEDLINE=99069613; PubMed=9851916;
RA      None;
RT      "Genome sequence of the nematode C. elegans: a platform for
RT      investigating biology. The C. elegans Sequencing Consortium.";
RL      Science 282:2012-2018(1998).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=BRISTOL N2;
RA      Tin-Wollam A., Henkhaus J., Wohldmann P., Osborn A.;
RT      "The sequence of C. elegans cosmid K04F1.";
RL      Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RN      [3]
RP      SEQUENCE FROM N.A.
RC      STRAIN=BRISTOL N2;
RA      Waterston R.;
RT      "Direct Submission.";
RL      Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AF106575; AAC78174.2; -.
DR      HSSP; P05091; 1CW3.
DR      InterPro; IPR002086; Aldehyde_dehydr.
DR      Pfam; PF00171; aldedh; 1.
DR      PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR      PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW      Hypothetical protein.

```

```

SQ      SEQUENCE   537 AA;  58601 MW;  F88875C8B460003D CRC64;

Query Match          33.5%; Score 853.5; DB 5; Length 537;
Best Local Similarity 41.7%; pred. No. 5.4e-58;
Matches 203; Conservative 75; Mismatches 184; Indels 25; Gaps 11;

QY    11 ELFINNEFVSSKSERLTLTNPWDESTVATDVHVANAADVDSAVAASQAVAKG-PWKKF 69
       :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    58 QLFINNEFVDKSGKTFFFNVPANGKLLA-KVAEGNRDDVDIAVEAAKKAFKIGSEWRRM 116
       : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY    70 TGAQRAACMLKFPADLAEKNAEKLARLES LPTGRPVSMITHFDIPNMVSVFRIYAGWADKI 129
       : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    117 DASHRGVLLNRLADLMERDRVILASLESLDNKGPKYEAYNIDLPSIKTFRIYAGYADKN 176
       : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY    130 AGKTFPEDNGKPWW-RYPEMGVCAGIASWNATFLVGVWKIAPALAAAGCSFIKASEKSPL 188
       || | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    177 HGKTI PVGGDYFTTRHEPVGCGQII PNFP LLMQAWKLAPALAMGNTVVNMKVAVKTPL 236
       : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY    189 GVLGLAPLFAEAGPPGVVQFLTGARV -TGEALASHMDIAKISFTRSVGGGGRAVKQATLK 247
       | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | :
Db    237 SALHVASLIKEAQPEGVVNIIIPGRGT DAGEAIASHMDVDVKVAF TGSTEVGKTIMKAAAE 296
       | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | :

QY    248 SNMKRVTTLELGEK-PTIVFNEAPLE---RQGESA KDFSKFGQIWVPPSCLLVQWGNLAE 303
       || : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | :
Db    297 SNVKKVTLELGKSPNIVFADADLEEAVRQSHHAL--FFNQGCCSAGSRRTFVE-GKIYD 353
       | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | :

QY    304 KFHG-----VRHGFEGGCORWLGNONPLEPKRTHGPFDVKSOYDRV LGNIDVG-KDTAQLL 357
       : L : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | :
Db    354 EFVAKAKELVEKTVIG-----DPFDENTTOGPIDESQVETIMKYIESGKKEGAQLV 405
       : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | :

QY    358 TGVGRKGDKGFAIEPTIFVNP KP GSKIWFEEIFGPVLSIKTKTEEAIEIANDTTYGLA 417
       || : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | :
Db    406 TGGVKHGDQGYFYFKPTIFANVNDQM KIAQE E I FGPNVI V IRFDSMEELIEKANNTTYGLA 465
       : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | :

QY    418 SVIYTKSLNRGLRVSSALETGGSINFPPIPETQT PFGGMKQSGSGRELGEGLKAYLEP 477
       : : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | :
Db    466 AGVVTNDLNKALQVANTIRAGSVWVNCYDVFDPAAPFGGFKQSGIGRELGEYLAAYTEV 525
       : : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | :

QY    478 KTNIHV 484
       || : | |

Db    526 KTVTIRV 532

```

RESULT	9				
Q90XS8					
ID	Q90XS8	PRELIMINARY;	PRT;	518	AA.
AC	Q90XS8;				
DT	01-DEC-2001	(TREMBlrel. 19,	Created)		
DT	01-DEC-2001	(TREMBlrel. 19,	Last sequence update)		
DT	01-DEC-2001	(TREMBlrel. 19,	Last annotation update)		
DE	RETINALDEHYDE DEHYDROGENASE TYPE 2.				
GN	RALDH2.				
OS	Brachydanio rerio (zebrafish) (zebra danio).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eu				
OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei; O				
OC	Cypriniformes; Cyprinidae; Danio.				
OX	NCBI_TaxID=7955;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=21543521; PubMed=11688558;				
RA	Begemann G., Schilling T.F., Rauch G.-J., Geisler R.,				
RT	"The zebrafish neckless mutation reveals a requirement				
RT	mesodermal signals that pattern the hindbrain.";				
RL	Development 128:3081-3094(2001).				
DR	EMBL; AF339837; AAL00899.1; -.				
SO	SEQUENCE 518 AA; 56515 MW; 2C6C451617E8E086 CRC64;				

Query Match 33.3%; Score 849.5; DB 13; Length 518;
Best Local Similarity 41.8%; Pred. No. le-57;
Matches 205; Conservative 78; Mismatches 195; Indels 13; Gaps 11;

Db	143	GQTIETDIKRFAYTRHEPIGVCQIIIPWNEPFLMCAWKIAPAVACGNTIILKTAELTPLS	202
Qy	190	VLGLAPLFAEAGFPPGVQFLTG-ARVTGEALASHMDIAKISFTRSVGGGRAVKQATLKS	248
Db	203	ALCLTKFVPECGFPPGVINVLSGDGRRCGNAISSHMDIDKVAFTGSTGVGRVMRAAASS	262
Qy	249	NMKRVTLLEGEK-PTIVFNEAPLERQSGESAKDFSKE-----QOIWVPESCLLVQWGNL	301
Db	263	NLKKVTLELGGKSPNIVFNDADL-----DSAAVWTNYGIFYNSGOVCCAGSRVYVQ-EDV	316
Qy	302	AEKFHGVHRHSGFCGQORWLGQ-----NLEPKRTHGPFVDKSOYDRVLGNIDVG-KDT	353
Db	317	YDEF-----IKRMVAKAKTLKVGDPFAEDTFQGAQVSKQOYERIVSVIESGIAHG	366
Qy	354	AQLLTGVGRKDGKGFAPTEPTIFVNPKP GSKIWFEETFGPVLSIKTFKTEEPAIEIANDTT	413
Db	367	AKLEIGGKRHGNLGYFVEPTILSNVTEDMAVGKEEIFGVPVLAVIKFKTIEAIRRGNNST	426
Qy	414	YGLASVIYTKSLNRGLRVSSALETGGVSNPPPIPETQTPEGGMKQSGSGRELGEGLKA	473
Db	427	YGLAAGVHTNITNAIKVSNALEAGTVWVNCYNLLHHQIPEGGYKESGIGRELGSYGLTN	486
Qy	474	YLEPKTINIHVNIE	487
Db	487	YTQTKAVHINLGMD	500

Search completed: April 27, 2002, 07:57:44
Job time: 700 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2002, 07:59:37 ; Search time 35.22 Seconds
(without alignments)
535.390 Million cell updates/sec

Title: US-09-882-694A-5
Perfect score: 2549
Sequence: 1 MVLSPEYKSELFINNERVS.....EEGLKAYLEPKTINHVNIE 487

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	882.5	34.6	517	1	DHA5_HUMAN
2	866	34.0	496	1	DHAL_CLAHE
3	864	33.9	497	1	DHAL_EBENI
4	855.5	33.6	500	1	DHAL_HORSE
5	855.5	33.6	500	1	DHAL_MOUSE
6	855.5	33.6	500	1	DHAL_RAT
7	848.5	33.3	500	1	DHAL_RAT
8	846.5	33.2	499	1	DHA2_HUMAN
9	843.5	33.1	500	1	DHAL_HUMAN
10	842	33.0	497	1	DHAL_ASPNG
11	840.5	33.0	499	1	DHAS_CHICK
12	840.5	33.0	517	1	DHAM_HUMAN
13	838.5	32.9	501	1	DHAE_MACPR
14	838.5	32.9	509	1	DHAL_CHICK
15	836.5	32.8	499	1	DHA2_MOUSE
16	835.5	32.8	499	1	DHA2_RAT
17	834.5	32.7	519	1	DHAM_MOUSE
18	833.5	32.7	519	1	DHAM_RAT
19	832.5	32.7	501	1	DHAE_ELEED
20	830.5	32.6	500	1	DHAL_SHEEP
21	829.5	32.5	520	1	DHAM_BOVIN
22	826.5	32.4	500	1	DHAL_BOVIN
23	826.5	32.4	500	1	DHAM_MESAU
24	824.5	32.3	512	1	DHA6_HUMAN
25	815.5	32.0	495	1	DHAL_ALTAL
26	810.5	31.8	497	1	DHAL_ENCBU
27	808.5	31.7	500	1	DHAM_HORSE
28	775	30.4	500	1	DHAL_AGABI
29	772.5	30.3	495	1	CROM_OCTDO
30	768	30.1	494	1	CROM_OMMSL
31	727	28.5	498	1	DHAM_LEITA
32	724.5	28.4	495	1	DHAL_ECOLI
33	723.5	28.4	490	1	DHAM_BACSU

34	715	28.1	519	1	DHA5_YEAST	P40047 saccharomyc
35	707.5	27.8	501	1	DHAB_AMAHP	O04895 amaranthus
36	707.5	27.8	506	1	DHA2_YEAST	P47771 saccharomyc
37	707	27.7	902	1	FTDH_HUMAN	Q75891 homo sapien
38	703	27.6	902	1	FTDH_RAT	P28037 rattus norv
39	702	27.5	497	1	DHAB_SPIOL	P17202 spinacia ol
40	699.5	27.4	499	1	FEAB_ECOLI	P80668 escherichia
41	696.5	27.3	500	1	DHA6_YEAST	P54115 saccharomyc
42	693	27.2	519	1	DHA4_YEAST	P46367 saccharomyc
43	692.5	27.2	500	1	DHAB_BETVU	P28237 beta vulgar
44	690.5	27.1	506	1	DHA3_YEAST	P54114 saccharomyc
45	675.5	26.5	502	1	DHAB_ATRHO	P42757 atriplex ho

ALIGNMENTS

RESULT 1

DHA5_HUMAN

ID DHA5_HUMAN STANDARD; PRT; 517 AA.

AC P30837;

DT 01-JUL-1993 (Rel. 26, Created)

DT 01-JUL-1993 (Rel. 26, Last sequence update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)

DE Aldehyde dehydrogenase X, mitochondrial precursor (EC 1.2.1.3) (ALDH class 2).

GN ALDH1B1 OR ALDH5 OR ALDHX.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Testis;

RX MEDLINE=91286241; PubMed=2061311;

RA Hsu L.C., Chang W.-C.;

RT "Cloning and characterization of a new functional human aldehyde dehydrogenase gene."

RT J. Biol. Chem. 266:12257-12265(1991).

RN [2]

RP VARIANTS VAL-86 AND ARG-107.

RX MEDLINE=94063858; PubMed=8244338;

RA Sherman D., Dave V., Hsu L.C., Peters T.J., Yoshida A.;

RT "Diverse polymorphism within a short coding region of the human aldehyde dehydrogenase-5 (ALDH5) gene."

RL Hum. Genet. 92:477-480(1993).

CC 1- FUNCTION: ALDHS PLAY A MAJOR ROLE IN THE DETOXIFICATION OF ALCOHOL-DERIVED ACETALDEHYDE. THEY ARE INVOLVED IN THE METABOLISM OF CORTICOSTEROIDS, BIOGENIC AMINES, NEUROTRANSMITTERS, AND LIPID PEROXIDATION.

CC 1- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O -> an acid + NADH.

CC 1- PATHWAY: SECOND STEP IN ETHANOL UTILIZATION.

CC 1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).

CC 1- SUBCELLULAR LOCATION: Mitochondrial matrix.

CC 1- TISSUE SPECIFICITY: LIVER, TESTIS AND IN A LESSER EXTENT IN BRAIN.

CC 1- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

CC EMBL; M63967; AAA96830.1; -.

CC HSSP; P20000; IAG8.

CC MIM; 100670; -.

CC InterPro; IPR002086; Aldehyde_dehydr.

CC Pfam; PF00171; aldedh; 1.

CC PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.

CC PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.

CC Oxidoreductase; NAD; Mitochondrion; Transit peptide; Polymorphism.

KW

FT TRANSIT 1 17 MITOCHONDRION (POTENTIAL).
FT CHAIN 18 517 ALDEHYDE DEHYDROGENASE X.
FT NP_BIND 262 267 NAD (ADP PART) (BY SIMILARITY).
FT ACT_SITE 285 285 POTENTIAL.
FT ACT_SITE 319 319 POTENTIAL.
FT VARIANT 86 86 A -> V (IN ALDH1B1*2).
FT VARIANT 107 107 /FTid=VAR_002257.
FT VARIANT 107 107 L -> R (IN ALDH1B1*3).
FT SEQUENCE 517 AA; 57217 MW; 78AC173AAC687165 CRC64;
Query Match 34.6%; Score 882.5; DB 1; Length 517;
Best Local Similarity 41.9%; Pred. NO. 1.6e-61;
Matches 210; Conservative 80; Mismatches 176; Indels 35; Gaps 13;
QY 2 VLSPEYKSELFINNEFVSSKGSERLTLTNPWDESTVATDVHVA--NAADVDSAVAASVQ 59
Db 29 ILNPDIPYNQLFINNEWODAVSKKTEPTVNPTTGEVIG--HVAEGDRADVRAVKAARE 85
QY 60 AVKKG-PWKKFTGAQRAACMLKFADLAENAEKLAERLESPTGRPVSMITHFDIPNMVSV 118
Db 86 AFRGSPWRMDASERGLNLLADLVERDRVYLASLETLDNGRPFQESYALDDEVYKV 145
QY 119 FRYAGWADKIAGKTFPEDNGKP--NWRYPDMGVCAIASWNTFLYVGVKIAAPALAGC 176
Db 146 YRYFAGWADKWHGKTIPM-HGQHECFTRHEPVGVCGQIIPNPLVQGVKLAAPALATGN 204
QY 177 SFIFKASEKSPGLVGLAPLFAEAGFPVGVQELTGARVT-GEALASHMDIAKISFTRSV 235
Db 205 TVVMKVAEQTPLSALYLASLIKEAGFPVGVNIIITGYGPTAGAAIAQHMDVDKVAFTGST 264
QY 236 GGGRAVKQATLKSNNMKRVTLLEGEK-PTIVNEAPLE--RQGESAKDFSKFGQIWWPPP 291
Db 265 EVGHLIQKAAGDSNLKRVTLLEGGKSPSIVLADADMEHAVEQCHEAL--FFNMGQCCAG 322
QY 292 SCLELVQWG-----NLAKEFHGVHSGFCGQRLWQGNPLEKRTHTGPFVVDKSYDRVL 344
Db 323 SRTEVEESIYNEFLERTVEK-----AKQKVG-NPFELDTQQGPQVDKEQFERVL 371
QY 345 GNIDVG-KDTAQLTGVRKGDGKFAIEPTIFVNPKPGSKIWPFEIFGPVLSIKTFKTEE 403
Db 372 GYQLGQKEGAKLLCGGRFGRGFFIKPTVFGGVQDDMRIAKEEIEFGVPQPLFKFKIE 431
QY 404 EATEIANDTTYGLASVIYTKSLNRLRVSSALETGGVSINFPFIPETQTPFGGMKQSGSG 463
Db 432 EVVERANTRYGLAAAVFTROLDKAMYFTQALQAGTVVWNTYNTVNTCHTTPFGGKESGNG 491
QY 464 RELGEEGLKAYLEPKTINIHV 484
Db 492 RELGEDGLKAYTEVKTIVK 512

RESULT 2
DHALLCLAEH STANDARD; PRT; 496 AA.
AC P40108;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Aldehyde dehydrogenase (EC 1.2.1.3) (ALDDH) (Allergen Cla h 3) (Cla h III).
GN CLAH3.
OS cladosporium herbarum.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina;
OC Dothidiomycetes et Chaetothyriomycetes incertae sedis;
OC Mycosphaerellaceae; mitosporic Mycosphaerellaceae; Cladosporium.
OX NCBI_TaxID=29918;
RN 1;
RP SEQUENCE FROM N.A.
RC STRAIN=280202-BERLIN;
RX MEDLINE=95206305; PubMed=7898496;
RA Achatz G., Oberkofler H., Lechenauer E., Simon B., Unger A.,

RA Kandler D., Ebner C., Prillinger H., Kraft D., Breitenbach M.;
RT "Molecular cloning of major and minor allergens of Alternaria
RT alternata and Cladosporium herbarum.";
RL Mol. Immunol. 32:213-227(1995).
CC -I- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O -> an acid + NADH.
CC -I- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -I- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X78228; CAA55072.1; -.
DR HSSP: P05091; 1CW3.
DR InterPro: IPR002086; Aldehyde_dehydr.
DR Pfam: PF00171; aldedh; 1.
DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW Oxidoreductase; NAD; Allergen.
FT ACT_SITE 263 263 BY SIMILARITY.
FT ACT_SITE 296 296 BY SIMILARITY.
SQ SEQUENCE 496 AA; 53397 MW; 4337C6B90E167990E CRC64;

Query Match 34.0%; Score 866; DB 1; Length 496;
Best Local Similarity 42.8%; Pred. NO. 3e-60;
Matches 211; Conservative 67; Mismatches 183; Indels 32; Gaps 12;
QY 7 EYKSELFINNEFVSSKGSERLTLTNPWDESTVATDVHVAADVDSAVAASVQAVKKGPM 66
Db 15 EQPTGLFINNEFVKGEGKTFDVINPSDES-VITQVHEATEKDVDIAVAARQAF-EGSW 72
QY 67 KKFTGAQRAACMLKFADLAENAEKLAERLESPTGRPVSMITHFDIPNMVSV-----FRY 121
Db 73 RLETPENRGKLLNNLANLFEKNTDLLAAVESLDNGKATSM-----ARVTSACASGCLRY 126
QY 122 YAGWADKIAGKTFPEDNGKPNMRY--EPMGVCACTASWNTFLYVGVKIAAPALAGCSF 178
Db 127 YGGWADKITGKVI--DTTPDTFNVYKKEPIGVCRSDHSLELPLLMWANKIGPAICGNTV 184
QY 179 IFKASEKSPGLVGLAPLFAEAGFPVGVQELTG-ARVTGEALASHMDIAKISFTRSVGG 237
Db 185 VLKTAEDTPLGGLVAASLVKEAGFPVGVINVISGFGKAVAGAAALSSHMDVDKVAFTGSTVV 244
QY 238 GRAVKQATLKSNNMKRVTLLEGEK-PTIVNEAPLERQ-SGESAKDFSKFGQIWWPPSCLL 295
Db 245 GRTILKAAASNNLKKVTLLEGGKSPNIVFEDADIDNAISWNVNFGIFFNHGQCCAGSRVY 304
QY 296 VQ---WGNLAKEFHGVHSGFCGQRLWQGNPLEKRTHTGPFVVDKSYDRVLGNIDVGKD 352
Db 305 VQESIYDKFVQKFK-----ERAQKNVVGDPFAADTFQGPQVSKVQFDRIMEYIQAGKD 357
QY 353 T-AQLLTGVGRKGDGKFAIEPTIFVNPKPGSKIWPFEIFGPVLSIKTFKTEEEAIEIAND 411
Db 358 AGATVETGSGRKGDGKGYFIEPTIFSNVTEDMKIVKEEIFGPVCSIAKPKTKEDAIAKLGNA 417
QY 412 TTYGLASVIYTKSLNRLRVSSALETGGVSINFPFIPETQTPFGGMKQSGSGRELGEGL 471
Db 418 STYGLAAAVHTKLNLTAEVSNALKAGTVVWNTYNTLHHQMPFGGYKESGIGRELGEDAL 477
QY 472 KAYLEPKTINIHV 484
Db 478 ANYTQTKTVSIRL 490
RESULT 3
DHALLCLAEH STANDARD; PRT; 497 AA.
ID DHALLCLAEH
AC P08157;

DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Aldehyde dehydrogenase (EC 1.2.1.3) (ALDDH).
GN ALDA OR ASPA
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=5072;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87248080; PubMed=3036652;
RA Pickett M., Gwynne D.I., Buxton F.P., Elliott R., Davies R.W.,
RA Lockington R.A., Scazzocchio C., Sealy-Lewis H.M.;
RT *Cloning and characterization of the alda gene of Aspergillus
RT nidulans.*;
RL Gene 51:217-226(1987).
CC -|- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O -> an acid + NADH.
CC -|- PATHWAY: SECOND STEP IN ETHANOL UTILIZATION.
CC -|- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M16197; AAA33293.1; -.
DR PIR: A29055; A29055.
DR HSSP: P05091; ICW3.
DR InterPro: IPR002086; Aldehyde_dehydr.
DR Pfam: PF00171; aldedh; 1.
DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW Oxidoreductase; NAD.
FT NP_BIND 241 246 NAD (ADP PART) (BY SIMILARITY).
FT ACT_SITE 264 264 BY SIMILARITY.
FT ACT_SITE 298 298 BY SIMILARITY.
SQ SEQUENCE 497 AA; 54088 MW; 1711FCEA993E571B CRC64;

Query Match 33.9%; Score 864; DB 1; Length 497;
Best Local Similarity 42.3%; Pred. No. 4.4e-60;
Matches 203; Conservative 79; Mismatches 176; Indels 22; Gaps 11;

QY 12 LFINNEFVSSKSGSERLTITNPWDESTVATDVHVAADVDVAASVQAVKGPWKFTG 71
| | | | | : : : : : | | | | | : : : : : | | | | : : : : : |
Db 21 LFINNEFVKGVEKTFQVINPSNEK-VITSVHEATEKDVDAVAAA-RAAFEGPWRQVTP 78
| | | | | : : : : : | | | | | : : : : : | | | | : : : : : |

QY 72 AQRAACMLKFADLAENAEKARLESPTGRPVSMITHFDIPNMVSVFRYYAGWADKIAG 131
: : | | | : : : : : | | | | : : : : : | | | | : : : : : |
Db 79 SERGILINKLADLMERDIDTLAATESLDNGKAFM-AKVDLANSIGCLRYAYAGWADKIHG 137
| | | | | : : : : : | | | | | : : : : : | | | | : : : : : |

QY 132 KTFPEDNGKPNW-RYEPMGVCAGIASWNATFLYVGWGIAPALAAAGCSFIFKASEKSP 190
: | : : : : : | | | | | : : : : : | | | | : : : : : |
Db 138 QTIDTNPETLTYTRHEPVGCGQIIPWNPFLMWSWKIGPAVAAGNTVVLKTAQQTPLSA 197
| | | | | : : : : : | | | | | : : : : : | | | | : : : : : |

QY 191 LGLAPLFAEAGFPVGVQFLTG-ARVTGEALASHMDIAKISFTRSVGGGRVAKVQATL 249
| | | | | : : : : : | | | | | : : : : : | | | | : : : : : |
Db 198 LYAAKLIKEAPPAGVINVISGFGRTAGAAISSHMDIDKVAFTGSTLVGPTILQAAKSN 257
| | | | | : : : : : | | | | | : : : : : | | | | : : : : : |

QY 250 MKRVTLLELGEK-PTIVNEAPLERQ-SGESAKDFSKFGQIWWPPSCLLVQWGNLAEKF 305
: : | | | | : : : : : | | | | : : : : : | | | | : : : : : |
Db 258 LKKVTLLELGGKSPNIVFDDADIDNAISWANFEGIFFNHGQCCAGSRILVQEG-IYDKFVA 316
| | | | | : : : : : | | | | | : : : : : | | | | : : : : : |

QY 306 ---HGVRHGSGCQRLWQGNPLEKRTHTGPFVDKSDYDRVLGNIDVGKDT-AQLITGVG 361
: : | | | | : : : : : | | | | : : : : : | | | | : : : : : |
Db 317 RFKERAQKNKG-----NPFEQDTFGQPVQSQLQFDRIMEYINHGKKAATVATGGD 368
| | | | | : : : : : | | | | | : : : : : | | | | : : : : : |

QY 362 RKGDKGFAIEPTIFVNPKPGSKIWFEEIFGPNLSIKTFKTEEEAIEIANDTYGLASVIY 421
| | : : : : : | | : : : : : | | : : : : : | | : : : : : | : : : : : |

Db 369 RHGNECYFIQPTVFTDVTSDMKIAQEEIFGPPVTIOKFKDVAEAIKIGNSTDYGLAAAVH 428
QY 422 TKSINRGLRVSSALETGVSVINFPPIPETQTQTFPGMGKQSGSRELGEGLKAYLEPKTN 481
| | : : : : : | | : : : : : | | : : : : : | | : : : : : |
Db 429 TRNVNTAIRVSNALKAGTWINNMYNMYISYQAPFGGFKQSGLRELGSYALENYTOIKTVH 488

RESULT 4
DHAL_HORSE
ID DHAL_HORSE STANDARD; PRT; 500 AA.
AC P15437;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Aldehyde dehydrogenase 1A1 (EC 1.2.1.3) (Aldehyde dehydrogenase,
DE cytosolic) (ALDH class 1) (ALDH1) (ALDH-E1).
GN ALDH1A1 OR ALDH1.
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9796;
RN [1]
RP SEQUENCE
RX MEDLINE=84208025; PubMed=5723662;
RA von Bahr-Lindstrom H., Hempel J., Joernvall H.;
RT *The cytoplasmic isoenzyme of horse liver aldehyde dehydrogenase.
RT Relationship to the corresponding human isoenzyme.*;
RL Eur. J. Biochem. 141:37-42(1984).
CC -|- FUNCTION: BINDS FREE RETINAL AND CELLULAR RETINOL-BINDING PROTEIN-
CC BOUND RETINAL. CAN CONVERT/OXIDIZES RETINALDEHYDE TO RETINOIC ACID
CC (BY SIMILARITY).
CC -|- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O -> an acid + NADH.
CC -|- PATHWAY: SECOND STEP IN ETHANOL UTILIZATION.
CC -|- SUBUNIT: HOMOTETRAMER.
CC -|- SUBCELLULAR LOCATION: Cytoplasmic.
CC -|- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
DR PIR: S02302; S02302.
DR HSSP: P51977; 1BXS.
DR InterPro: IPR002086; Aldehyde_dehydr.
DR Pfam: PF00171; aldedh; 1.
DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW Oxidoreductase; NAD; Acetylation.
FT MOD_RES 1 1 ACETYLATION.
FT NP_BIND 245 250 NAD (ADP PART) (BY SIMILARITY).
FT ACT_SITE 268 268 BY SIMILARITY.
FT ACT_SITE 302 302 BY SIMILARITY.
SQ SEQUENCE 500 AA; 54743 MW; A438E4204BEF7340 CRC64;

Query Match 33.6%; Score 855.5; DB 1; Length 500;
Best Local Similarity 41.9%; Pred. No. 2e-59;
Matches 202; Conservative 87; Mismatches 182; Indels 11; Gaps 10;

QY 10 SELFNNNEFVSSKSGSERLTITNPWDESTVATDVHVAADVDVAASVQAVKKG-PWKK 68
: : | | | | : : : : : | | | | : : : : : | | | | : : : : : |
Db 20 TKIFINNEWHDSVSGKRPVPVFNPAATEKLC-EVEEGDKEDVKNKAVAAARQAFQIGSPWRT 78
| | | | | : : : : : | | | | | : : : : : | | | | : : : : : |

QY 69 FTCAQRAACMLKFADLAENAEKARLESPTGRPVSMITHFDIPNMVSVFRYYAGWADK 128
: : | | | | : : : : : | | | | : : : : : | | | | : : : : : |
Db 79 MDASERGRLLYKLADLVDRDLILATMESMNGKLFNSAYLMDLGGCLKTLRYCAGWADK 138
| | | | | : : : : : | | | | | : : : : : | | | | : : : : : |

QY 129 IAGKTFFPED-NGKPNWRYEPMGVCAGIASWNATFLYVGWGIAPALAAAGCSFIFKASEKSP 187
| | | | | : : : : : | | | | : : : : : | | | | : : : : : |
Db 139 IQGRTIPSDGNFFTYTRHEPVGCGQIIPWNPFLMFLMWKIAPALSCGNTVVVKPAEQTP 198
| | | | | : : : : : | | | | | : : : : : | | | | : : : : : |

QY 188 LGVIGLAPLFAEAGFPVGVQFLTGARVT-GEALASHMDIAKISFTRSVGGGRVAKVQATL 246
| | | | | : : : : : | | | | | : : : : : | | | | : : : : : |
Db 199 LSALHVAATLIKEAGFPVGVNIVPGYGPAGTAAGAAISSHMDIDKVAFTGSTEVGKLIKEAAG 258
| | | | | : : : : : | | | | | : : : : : | | | | : : : : : |

QY 247 KSNMKRVTLLELGEK-PTIVNEAPLERQSGSAGD-FSKFGQIWWPPSCLLVQWGNLAEK 304
| | : : : : : | | : : : : : | | : : : : : | | : : : : : | : : : : : |
Db 259 KSNLKRVTLELGGKSPFIVFADADLETALEVTHQALFYHQGCCVAAASRLFVE-ESIYDE 317

RL Alcohol 2:103-106(1985).
RN [9]
RP SEQUENCE OF 214-500 FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=87279033; PubMed=3610592;
RA Agarwal D.P., Goedde H.W.;
RT "Human aldehyde dehydrogenase isozymes and alcohol sensitivity.*";
RL Isozymes Curr. Top. Biol. Med. Res. 16:21-48(1987).
RN [10]
RP VARIANT LYS-504.
RX MEDLINE=84119449; PubMed=6582480;
RA Yoshida A., Huang I.-Y., Ikawa M.;
RT "Molecular abnormality of an inactive aldehyde dehydrogenase variant commonly found in Orientals.*";
RL Proc. Natl. Acad. Sci. U.S.A. 81:258-261(1984).
RN [11]
RP VARIANT LYS-496.
RX MEDLINE=96119362; PubMed=8561277;
RA Novoradovsky A., Tsai S.J., Goldfarb L., Peterson R., Long J.C., Goldman D.;
RT "Mitochondrial aldehyde dehydrogenase polymorphism in Asian and American Indian populations: detection of new ALDH2 alleles.*";
RL Alcohol. Clin. Exp. Res. 19:1105-1110(1995).
RN [12]
RP X-RAY CRYSTALLOGRAPHY (2.58 ANGSTROMS).
RX MEDLINE=20095857; PubMed=10631996;
RA Ni L., Zhou J., Hurley T.D., Weiner H.;
RT "Human liver mitochondrial aldehyde dehydrogenase: three-dimensional structure and the restoration of solubility and activity of chimeric forms.*";
RL Protein Sci. 8:2784-2790(1999).
CC -!- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O = an acid + NADH.
CC -!- PATHWAY: SECOND STEP IN ETHANOL UTILIZATION.
CC -!- SUBUNIT: HOMOTETRAMER.
CC -!- SUBCELLULAR LOCATION: Mitochondrial matrix.
CC -!- DISEASE: A VERY HIGH INCIDENCE OF ACUTE ALCOHOL INTOXICATION IN ORIENTALS AND SOUTH AMERICAN INDIANS, AS COMPARED TO CAUCASIANS, CAN BE DIRECTLY ATTRIBUTED TO AN ENZYMATICALLY IMPAIRED ALDH ISOZYME. THE INACTIVE VARIANT (ALLELE 2 OR ALDH2*2) IS DUE TO SINGLE AMINO ACID EXCHANGE.
CC -!- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
CC -!- CAUTION: THE CONFLICTS BETWEEN THE SEQUENCE DETERMINED IN REF.1 AND REF.2 AND THAT IN REF.3 AND 4 ARE PROBABLY ALL DUE TO FRAMESHIFT OR SEQUENCING ERRORS AS DESCRIBED IN REF.5 AND REF.6.
CC THE SEQUENCE DESCRIBED IN REF.9 DIFFERS FROM THAT SHOWN DUE TO TWO FRAMESHIFTS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; M20456; AAA51693.1; -.
DR EMBL; M20444; AAA51693.1; JOINED.
DR EMBL; M20445; AAA51693.1; JOINED.
DR EMBL; M20446; AAA51693.1; JOINED.
DR EMBL; M20447; AAA51693.1; JOINED.
DR EMBL; M20448; AAA51693.1; JOINED.
DR EMBL; M20449; AAA51693.1; JOINED.
DR EMBL; M20450; AAA51693.1; JOINED.
DR EMBL; M20451; AAA51693.1; JOINED.
DR EMBL; M20452; AAA51693.1; JOINED.
DR EMBL; M20453; AAA51693.1; JOINED.
DR EMBL; M20454; AAA51693.1; JOINED.
DR EMBL; X05409; CAA28990.1; -.
DR EMBL; Y00109; CAA68290.1; -.
DR EMBL; X03001; AAB59500.1; -.
DR EMBL; M26760; AAA51694.1; -.
DR EMBL; M54931; AAA62825.1; ALT_FRAME.
PIR; S00804; DEHUE2.

DR PIR; A29975; A29975.
DR PDB; 1CW3; 10-JAN-00.
DR MIM; 100650; -.
DR InterPro; IPR002086; Aldehyde_dehydr.
DR Pfam; PF00171; aldedh; 1.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW Oxidoreductase; NAD; Mitochondrion; Transist peptide; Polymorphism;
KW 3D-structure.
FT TRANSIT 1 17 MITOCHONDRION.
FT CHAIN 18 517 ALDEHYDE DEHYDROGENASE.
FT NP_BIND 262 267 NAD (ADP PART) (BY SIMILARITY).
FT ACT_SITE 285 285
FT ACT_SITE 319 319
FT VARIAT 337 337
FT VARIAT 496 496
FT VARIAT 504 504
FT VARIAT 504 504
FT CONFLICT 7 12 E -> V (IN DBSNP:1062136).
FT CONFLICT 7 26 /FTID=VAR_011869.
FT CONFLICT 18 18 E -> K (IN ALDH2*3).
FT CONFLICT 80 85 /FTID=VAR_011302.
FT CONFLICT 216 216 E -> K (IN ALDH2*2; DRASTIC REDUCTION
FT CONFLICT 218 218 OF ENZYME ACTIVITY).
FT CONFLICT 247 247 /FTID=VAR_002248.
FT CONFLICT 380 380 REFGLRL -> ARAPP (IN REF. 5).
FT CONFLICT 462 463 REFGLRL -> ARAPP (IN REF. 5).
SQ SEQUENCE 517 AA; 56381 MW; E8F74D44D285A00E CRC64;
Query Match 33.0%; Score 840.5; DB 1; Length 517;
Best local similarity 41.0%; Pred. No. 3.2e-58;
Matches 200; Conservative 86; Mismatches 187; Indels 15; Gaps 12;
QY 5 PDEYKSELFNNFEVSSKSGSERLTLTNPWDESTVATDVHVAADVDVAASVQAVKKG 64
DB 32 PEVFCNQIFINNEWHDAVSRKTFPTVNP-STGEVICQVAEGDKEDVDKAVKAARAEQLG 90
QY 65 -PWKFTGAQRAACMLKFAADLAENAKNAEKARLESFTGRPVSMITHFDIPNMVSVFRYYA 123
DB 91 SPWRMDASHGRLLNRLADLIERDRTYLALETLDNGKPYVISYLDLMDVLCCLRYA 150
QY 124 GWADKIAGKTFEPEDNGKPNW-RYEPMGVCAGIASNNATFLYGVWKIAPALAAAGCSFIFKA 182
DB 151 GWADKYHETIPIDGDFFSYTRHEPVGCGQIIPNFPPLMQAWKLGALATGNVVMKV 210
QY 183 SEKSPGLVLGLAPLFAEAGFPFGVQVQELTGARVT-GEALASHMDIAKISFTSVGGGRAV 241
DB 211 AEQTPLTALYVANLIKEAGFPFGVNVIVPGFGPTAGAAIASHEDVDKVAFTGSTEIGRVI 270
QY 242 KQATLKSNMKRVTLLEGEK-PTIVNEAPLERQSGESAKDFSKF---CQIWVPPSCLLVQ 297
DB 271 QVAAGSSNLKRVTLLEGGKSPNLIIMSDADM-NAVEQA-HFALFFNQGCCAGSRFTVQ 328
QY 298 WGNLAERFHVHSGFGGCGQWLGONPLEKPRTHGPFVDKQYDRVLGNIDVGK-DTAQL 356
DB 329 -EDIYDEF--VERSVARAKSRVVG-NPFDSTKEQGPQVDETQFKKILGYINTKQEGAKL 384
QY 357 LTGVGRKDGKGAIEPTIFVNPKPGSKIWFEEIFGPVLSIKTFKTEEAIEIANDTTYGL 416
DB 385 LCGGGIAADRGYFIQPTVFGVDQDGMTIAKEEIFGPVMOILKFKTIEVVGGRANNSTYGL 444
QY 417 ASVIYTKSLNRLRVSSALETGGVSINFPFIPETQTPFGGMKQSGSGRELGEGLKAYLE 476
DB 445 AAAVFTKOLDKANYLSQALQACTVWVNCYDVFAGACSPFGYKMSGSGRELGEYGLQAYTE 504
QY 477 PKTINIHV 484
DB 505 VKTIVTKV 512

```
RESULT 13
DHAEL_MACPR
ID DHAEL_MACPR STANDARD; PRT; 501 AA.
AC Q29490;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Aldehyde dehydrogenase, cytosolic 1 (EC 1.2.1.3) (ALDH class 1) (ETA-
crystallin).
OS Macrosclerites proboscideus (Short-eared elephant shrew).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Macroscelidea; Macroscelididae; Macroscelides.
OX NCBI_TaxID=29082;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Eye;
RX MEDLINE=96279083; PubMed=8663049;
RA Graham C., Hodin J., Wistow G.;
RT "A retinaldehyde dehydrogenase as a structural protein in a mammalian
eye lens. Gene recruitment of eta-crystallin."
RL J. Biol. Chem. 271:15623-15628(1996).
CC -!- FUNCTION: MAJOR COMPONENT OF THE EYE OF ELEPHANT SHREWS, WHICH IN
CC CONTRAST TO OTHER MAMMALS, POSSESSES BOTH A LENS- AND A NON-LENS
CC CLASS-1 ALDEHYDE DEHYDROGENASE 1. THIS EYE-SPECIFIC FORM IS A
CC STRUCTURAL PROTEIN OF THE LENS AND, IN OTHER PART OF THE EYE,
CC SERVES AS THE MAJOR FORM OF ALDH1. CAN CONVERT/OXIDIZES
CC RETINALDEHYDE TO RETINOIC ACID.
CC -!- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O = an acid + NADH.
CC -!- PATHWAY: SECOND STEP IN ETHANOL UTILIZATION (BY SIMILARITY).
CC -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- TISSUE SPECIFICITY: EYE-SPECIFIC, WITH VERY HIGH EXPRESSION IN THE
CC LENS.
CC -!- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U03906; AAC48588.1; -
DR HSSP; P51977; 1BXS.
DR InterPro; IPR002086; Aldehyde_dehydr.
DR Pfam; PF00171; aldedh; 1.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; FALSE_NEG.
KW Oxidoreductase; NAD.
FT NP_BIND 246 251 NAD (ADP PART) (BY SIMILARITY).
FT ACT_SITE 269 269 POTENTIAL.
FT ACT_SITE 303 303 POTENTIAL.
SQ SEQUENCE 501 AA; 54624 MW; 48C15C682665D0C0 CRC64;

Query Match 32.9%; Score 838.5; DB 1; Length 501;
Best Local Similarity 40.6%; Pred. No. 4.3e-58;
Matches 196; Conservative 92; Mismatches 182; Indels 13; Gaps 10;

QY 10 SELFNNFEVSSKGSERLTLTNPWDESTVATDVHVANAADVSAVAQVAVKKG-PWKK 68
Db :|||||: | : | | | : | | : | | : | | | | : | | : | | : | | :
21 TKLFNNNEHDSVSGKTFPVFNPAEEKIC-EVEEADKEDVDKAVKAAREAFQMGSPWRT 79

QY 69 FTGAORAAACMLKFADLAENKNAEKLARLESLETPGRPVSMITHFDIPNMVSVFRRYAGWADK 128
Db :| : | | | : | | | | : | | | : | : | : | | | | | | | | | |
80 MDASERGLIYKLADLIERDRLLATLESINAGKIFASAYIMLDYCIKVLRYCAGWADK 139

QY 129 IAGKTFPEDNGKPNW-RYEPMGVCAGIASWNATFELYVGWKIAPALAAAGCSFFKASKSP 187
Db | | | | : | : | | | | | | | | : | | | | : | | | | : | : | : |
140 IQGRTIPVDGEFFSYTRHEPIGVCGQIFPNWAPMILLACKIGPACCGNTVIVKPAEQTP 199
```

```
QY 188 LGVLGLAPLFAEAGFPFVVQFLTGARVT-GEALASHMDIAKISFTRSVGGGKAVKQATL 246
| | | | | | | | | | : | | | | | | | | | | : | | : | | :
Db 200 LTAHVASLIKEAGFPFVVNIVPGYGTAGAAISSHMDVDKVAFTGSTEVGKMIQRAAA 259
| | | | | | | | | | : | | | | | | | | | | : | | : | | :
QY 247 KSNMKRVTLLEGEK-PTIVFNEAPLERQSGESAKD--FSKFGQIWWPPSCLLVQWGNLAE 303
| | | | | | | | | | : | | | | | | | | | | : | | : | | :
Db 260 KSNLKRVTLELGAKNPCIVFADADLD-SAVEFAHQGVFTNQOGSCIAASKLFEV-ETIYD 317
| | | | | | | | | | : | | | | | | | | | | : | | : | | :
QY 304 KFHGVRHGSFGGCGQWLGNPLPKRTHGPFVDKSYDRVLGNIDVG-KDTAQLLTGVGR 362
| | | | | | | | | | : | | | | | | | | | | : | | : | | :
Db 318 EF---VQSRVERAKKYVFGNPLTPGVNHGPNQKQAQHNKIMELIESGKKEGAKLECGGP 374
| | | | | | | | | | : | | | | | | | | | | : | | : | | :
QY 363 KGDGGAIEPTIFVNPXPGSKIWFEEIFGPVLSIKTFKTEEEAIEIANDTTYGLASVIYT 422
| | | | | | | | | | : | | | | | | | | | | : | | : | | :
Db 375 WGNKGYFIQPTIFSNTDMRIAKEEIEFGPQOIMKFKSLDEVIKRANNTYYGLVAGVET 434
| | | | | | | | | | : | | | | | | | | | | : | | : | | :
QY 423 KSLNRGLRVSSALETGGVSYNFPFIPETOTPFPGMKQSGSGRELGEGLKAYLEPKTINI 482
| | | | | | | | | | : | | | | | | | | | | : | | : | | :
Db 435 KDLDAKAVTVSSALQAGTWWVNCYLAASAQSPAGGFKMSGHGREMGEYGIHEYTEVKTVTM 494
| | | | | | | | | | : | | | | | | | | | | : | | : | | :
QY 483 HVN 485
Db 495 KIS 497

RESULT 14
DHAEL_CHICK
ID DHAEL_CHICK STANDARD; PRT; 509 AA.
AC P27463;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Aldehyde dehydrogenase 1A1 (EC 1.2.1.3) (Aldehyde dehydrogenase,
cytosolic) (ALDH class 1).
GN ALDH1A1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryonic retina;
RX MEDLINE=92217647; PubMed=1559558;
RA Godbout R.;
RT "High levels of aldehyde dehydrogenase transcripts in the
undifferentiated chick retina."
RL Exp. Eye Res. 54:297-305(1992).
CC -!- FUNCTION: BINDS FREE RETINAL AND CELLULAR RETINALDEHYDE TO RETINOIC ACID
CC BOUND RETINAL. CAN CONVERT/OXIDIZES RETINALDEHYDE TO RETINOIC ACID
CC (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O = an acid + NADH.
CC -!- PATHWAY: SECOND STEP IN ETHANOL UTILIZATION.
CC -!- SUBUNIT: HOMOTETRAMER.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X58869; CAA41679.1; -
DR PIR; S14629; S14629.
DR HSSP; P51977; 1BXS.
DR InterPro; IPR002086; Aldehyde_dehydr.
DR Pfam; PF00171; aldedh; 1.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
```


KW Oxidoreductase: NAD.
FT NP_BIND 254 259 NAD (ADP PART) (BY SIMILARITY).
FT ACT_SITE 277 277 BY SIMILARITY.
FT ACT_SITE 311 311 BY SIMILARITY.
SQ SEQUENCE 509 AA; 55809 MW; 7771181PA2F05DA9 CRC64;

Query Match 32.9%; Score 838.5; DB 1; Length 509;
Best Local Similarity 39.8%; Pred. No. 4.4e-58;
Matches 192; Conservative 87; Mismatches 190; Indels 13; Gaps 8;

QY 10 SELFNNFVSSKGSERLTLTNPWDESTVATDVHANAADVDSDAVASQAVKKG-PWKK 68
Db 29 TKIFINNEWHDSVSGKKFVEFNPANEEKIC-EVAEGDKADIDKAVKARKAFELGSPWRT 87
QY 69 FTGAORAAACMLKFAADLAEKNAEKIARLESPTGRPVSMITHFDIPNMVSVFRYYAGWADK 128
Db 88 MDASERGRLLNKLADLVERDRITLATMEADIDGKIFSTAYLMDLGACIKTIRYCAGWADK 147
QY 129 IAGKTFPED-NGKPNWRYEPMGVCAIASWNAATFLYVGWKIAPALAAAGCSFIFKASEKSP 187
Db 148 IHGRTVPMGDGNFTFTTRHEPVGVCGQIIIPWNPFLVFIWIKIAPALCCGNTVVVKPAEQTP 207
QY 188 LGVLGLAPLFAEAGFPVGVQELTGARVT-GEALASHMDIAKISFTRSVGGGRAVKQATL 246
Db 208 LSALYMGSLIKEAGFPVGVNIVPGFGPTAGAAISHHMDIDKVSFTGSTEVGKLIKEAAG 267
QY 247 KSNMKRVTLLEGEK-PTIVNEAPLERQSGESAKDFSKFGQIWPVPSCLLVQWGNLAEK- 304
Db 268 KTNLKRVTLEGGKSPNIIIFADADL-----DEAEFAHIGLFYHGGQCCAGSRIFVEEP 322
QY 305 -FHGVRHSGSGGQCORWLGQNPLEPKRTHGPFVDKSDYRVLGNIDVG-KDTAQLLTGVGR 362
Db 323 IYDEFVRSIERAKKYTLGDPPLPCVQVQGPQIDKEQFQKILDLESKGKGAKECGGGP 382
QY 363 KGDKGPAIEPTIFVNPCKPGSKIWFEEIFGPPVLSIKTFKTEEEATEIANDTTYGLASVIYT 422
Db 383 WGNKGFLQPTVFSNVTDMDRIAKEEITGPPVQIQIMKFKTIDEVTKRANNTTYGLAAAVFT 442
QY 423 KSLNRLVRSSALETGGVSINFPFIPETQTPFGGMKOSGSGRELGEGLKAYLEPKTINI 482
Db 443 KDIDKALTFASALQAGTVWVNCYSAFSAQCDFGGFKMSGNGRELGEYGLQYETEVTVTI 502
QY 483 HV 484
Db 503 KI 504

RESULT 15
DHA2_MOUSE
ID DHA2_MOUSE STANDARD; PRT; 499 AA.
AC Q62148;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Aldehyde dehydrogenase 1A2 (EC 1.2.1.3) (Retinaldehyde-specific
dehydrogenase type 2) (RALDH(II)) (RALDH-2).
GN ALDH1A2 OR ALDH1A7 OR RALDH2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/HE;
RX MEDLINE=96390857; PubMed=8797830;
RA Zhao D., McCaffery P., Ivins K.J., Neve R.L., Hogan P., Chin W.W.,
RA Draeger U.C.;
RT "Molecular identification of a major retinoic-acid-synthesizing
RT enzyme, a retinaldehyde-specific dehydrogenase."
RL Eur. J. Biochem. 240:15-22(1996).
CC -!- FUNCTION: RECOGNIZES AS SUBSTRATES FREE RETINAL AND CELLULAR
CC RETINOL-BINDING PROTEIN-BOUND RETINAL. DOES METABOLIZE OCTANAL AND

CC DECANAL BUT DOES NOT METABOLIZE CITRAL, BENZALDEHYDE, ACETALDEHYDE
CC AND PROPANAL EFFICIENTLY (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O = an acid + NADH.
CC -!- PATHWAY: RETINOIC ACID BIOGENESIS.
CC -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X99273; CAA67666.1; -
DR HSSP; P51977; 1BX5.
DR MGD; MGI:107928; Aldh1a2.
DR InterPro; IPR002086; Aldehyde_dehydr.
DR Pfam; PF00171; aldedh; 1.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW Oxidoreductase; NAD.
FT NP_BIND 244 249 NAD (ADP PART) (BY SIMILARITY).
FT ACT_SITE 267 267 POTENTIAL.
FT ACT_SITE 301 301 POTENTIAL.
SQ SEQUENCE 499 AA; 54725 MW; E0000596A90B07A CRC64;

Query Match 32.8%; Score 836.5; DB 1; Length 499;
Best Local Similarity 40.7%; Pred. No. 6.2e-58;
Matches 196; Conservative 83; Mismatches 191; Indels 11; Gaps 10;

QY 10 SELFNNFVSSKGSERLTLTNPWDESTVATDVHANAADVDSDAVASQAVKKG-PWKK 68
Db 19 TKIFINNEWHDSVSGRVFVPCNPATGEQVC-EVQEDKVIDDKAVQAARLAFSLGSWRR 77
QY 69 FTGAORAAACMLKFAADLAEKNAEKIARLESPTGRPVSMITHFDIPNMVSVFRYYAGWADK 128
Db 78 MDASERGRLLNKLADLVERDRATLATMESLNGGKPFQAFYIDLOGVIKTLRYAGWADK 137
QY 129 IAGKTFPEDNGKPNW-RYEPMGVCAGIASWNAATFLYVGWKIAPALAAAGCSFIFKASEKSP 187
Db 138 IHGRTIPVDGDFYFTTRHEPIGVCGQIIPWNPFLMPTWKIAPALCCGNTVVIKPAEQTP 197
QY 188 LGVLGLAPLFAEAGFPVGVQELTGARVT-GEALASHMDIAKISFTRSVGGGRAVKQATL 246
Db 198 LSALYMGALIKEAGFPVGVNIIIPGYGPTAGAAIASHIGIDKIAFTGSTEVGKLIQEAAG 257
QY 247 KSNMKRVTLLEGEK-PTIVNEAPLERQSGESAKD-FSKFGQIWPVPSCLLVQWGNLAEK 304
Db 258 RSNLKRVTLEGGKSPNIIIFADADLDYAVEQAHOGVEFNQGCCTAGSRIFVE-ESIYEE 316
QY 305 FHGVRHSGSGGQCORWLGQNPLEPKRTHGPFVDKSDYRVLGNIDVG-KDTAQLLTGVGRK 363
Db 317 F--VKRSVERAKRRIVG-SPFDPTTEQGPQIDKKQYKNKVLIELIQSGVAEGAKLECGGKGL 373
QY 364 GDKGPAIEPTIFVNPCKPGSKIWFEEIFGPPVLSIKTFKTEEEAIEIANDTTYGLASVIYTK 423
Db 374 GRKGFFIEPTVFSNVTDMDRIAKEEITGPPVQIEILRFTKTMDEVIERANNSDFGLVAAVFTN 433
QY 424 SLNRLVRSSALETGGVSINFPFIPETQTPFGGMKOSGSGRELGEGLKAYLEPKTINI 483
Db 434 DINKALMVSSAMQAGTVWVNCYNALNAQSPFGFKMSGNGREMGEFGLREYSEVKTVTVK 493
QY 484 V 484
Db 494 I 494

Job time: 745 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2002, 07:48:20 ; Search time 68.15 Seconds
(without alignments)
686.655 Million cell updates/sec

Title: US-09-882-694A-5
Perfect score: 2549
Sequence: 1 MVLSPEYKSELFIFINFEVS.....EEGLKAYLEPKTINIHVNIE 487

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	889.5	34.9	517	1 A40872	aldehyde dehydroge
2	866	34.0	496	2 S43114	aldehyde dehydroge
3	864	33.9	497	1 A29055	aldehyde dehydroge
4	856.5	33.6	501	1 JQ1004	aldehyde dehydroge
5	855.5	33.6	500	1 S02302	aldehyde dehydroge
6	855.5	33.6	501	1 A32616	aldehyde dehydroge
7	848.5	33.3	501	2 JC5553	aldehyde dehydroge
8	843.5	33.1	501	1 DEHUE1	aldehyde dehydroge
9	843.5	33.1	501	2 JC4524	aldehyde dehydroge
10	840.5	33.0	517	1 DEHUE2	aldehyde dehydroge
11	838.5	32.9	509	1 S14629	aldehyde dehydroge
12	836.5	32.8	499	2 S74224	aldehyde dehydroge
13	834.5	32.7	519	1 I48966	aldehyde dehydroge
14	833.5	32.7	519	1 S03564	aldehyde dehydroge
15	830.5	32.6	501	2 S14752	aldehyde dehydroge
16	829.5	32.5	520	1 S09030	aldehyde dehydroge
17	824.5	32.3	512	1 A55684	aldehyde dehydroge
18	820	32.2	503	2 T39216	aldehyde dehydroge
19	815.5	32.0	495	2 S43108	aldehyde dehydroge
20	810.5	31.8	497	2 JC4924	aldehyde dehydroge
21	808.5	31.7	500	1 S00364	aldehyde dehydroge
22	792	31.1	544	2 D88449	protein F54D8.3 [1
23	781	30.6	538	2 T06683	rf2 nuclear restor
24	779	30.6	549	2 T03983	probable aldehyde
25	773.5	30.3	496	2 T50272	omega-crystallin -
26	772.5	30.3	496	1 A46725	aldehyde dehydroge
27	751	29.5	542	2 T02301	aldehyde dehydroge
28	737.5	28.9	495	2 H69614	aldehyde dehydroge
29	730.5	28.7	495	2 E90863	aldehyde dehydroge

30	730.5	28.7	495	2 D85755	aldehyde dehydroge
31	728.5	28.6	519	2 C86372	hypothetical prote
32	727	28.5	498	1 S43184	aldehyde dehydroge
33	725.5	28.5	498	2 C83717	NADP-dependent ald
34	724.5	28.4	495	1 G64878	aldehyde dehydroge
35	723.5	28.4	490	2 A69629	glycine betaine al
36	721.5	28.3	520	1 S50576	probable aldehyde
37	702	27.5	497	1 A35994	betaine-aldehyde d
38	700.5	27.5	500	2 T51172	betaine-aldehyde d
39	699.5	27.4	500	2 D64889	probable phenylace
40	699	27.4	497	2 T51173	betaine-aldehyde d
41	696.5	27.3	497	2 B82981	probable aldehyde
42	693	27.2	519	1 S67286	probable aldehyde
43	692.5	27.2	500	1 S19135	betaine-aldehyde d
44	691	27.1	500	1 S60929	probable aldehyde
45	690.5	27.1	506	1 S54527	aldehyde dehydroge

ALIGNMENTS

RESULT 1

A40872

aldehyde dehydrogenase (NAD+) (EC 1.2.1.3) 5 precursor, mitochondrial - human
C:Species: Homo sapiens (man)
C:Date: 27-Mar-1992 #sequence_revision 31-Jan-1997 #text_change 11-Jun-1999
C:Accession: A40872
R:Hsu, L.C.; Chang, W.C.
J. Biol. Chem. 266, 12257-12265, 1991
A:Title: Cloning and characterization of a new functional human aldehyde dehydrogenase
A:Reference number: A40872; MUID:91286241
A:Molecule type: DNA; mRNA
A:Residues: 1-517 <HSU>
A:Cross-references: GB:M63967; NID:g337184; PIDN:AAA96830.1; PID:g1263008
A:Note: 86-Ala and 107-Leu were also found
C:Comment: Based on similarity, this form is assumed to be mitochondrial and to have
C:Genetics:
A:Gene: GDB:ALDH5; ALDHX
A:Cross-references: GDB:128788
A:Map position: 9p13-9p13
A:Note: the only intron occurs before the initiator codon
C:Complex: homotetramer (probably)
C:Function:

A:Description: catalyzes oxidation of an aldehyde to an acid using NAD+ and water
A:Pathway: ethanol catabolism
A:Note: enzymes with this activity are involved in diverse metabolic pathways in var
C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
C:Keywords: alcohol metabolism; homotetramer; liver; mitochondrion; NAD; oxidoreduct
F:1-20/Domain: transit peptide (mitochondrion) #status predicted <TNP>
F:21-517/Product: aldehyde dehydrogenase (NAD+) 5 #status predicted <MAT>
F:75-339/Domain: aldehyde dehydrogenase homology <ALDD>
F:285,319/Active site: Glu, Cys #status predicted

Query Match 34.9%; Score 889.5; DB 1; Length 517;
Best Local Similarity 41.8%; Pred. No. 5.5e-62;
Matches 209; Conservative 80; Mismatches 178; Indels 33; Gaps 12;

QY	2	VLSPDEYKSELFIFINFEVSSKGSERITLTNPWDESTVAIDVHVA--NAADVDSAAVASVQ	59
DB	29	ILNPDIPYNQLFINNEWQDAVSKKTPPTVNPPTTGEVIG--HVAEGDRAVDRAVKAARE	85
QY	60	AVKKG-PWKFTGAORACMLKFAADLAENAEKARLESPLTGRPVSMITHFDIPNMVSV	118
DB	86	VFRGLSPWRMRDASERGLLNRLADLVERDRVYLASLETLDNGKPFQESYALDDEVIKV	145
QY	119	FRYAGWADKIAGKTFPEDNGKPNW-RYEPMGVCAGIASWNATFELYVGWKIAPALAGCS	177
DB	146	YRYFAGWADKWHGKTIPTMDGQHCFTRHEPVGVCQIIPWNEPLVMQCKLAPALATGNT	205
QY	178	FIFKASEKSPILGVLGLAPLFAEACFPFGVQFLTGARVT-GEALASHMDIAKISFTRSVG	236

Db 206 VVMKVAEOTPLSALYLASLIKEAGFPFGVWNIIITGYGPTAGAAIAQHMDVDKVAFTGST 265

QY 237 GGRAVKQATLKSNMKRVTLLELGEK-PTIVFNEAPLE---RQSGESAKDFSKFGQIWPVPS 292

Db 266 VGHLIQKAAGDNLKRVTLLELGGKSPSIVLADADMEHAVEQCHEAL--FFNMGQCCAGS 323

QY 293 CLLVQWG-----NLAEKPHGVRHGSFGGRCORWLQONPLEPKRTHGPFVDKSYDRVLG 345

Db 324 RTFVEESLYNEFLERTVEK-----AKQRKVG-NPFELDTQQGPPQVDREQFERVLG 372

QY 346 NIDVG-KDTAQLLTGVRGKDGKGFATIEPTIFVNPKPCKSIWFEEIFGPPVLSIKTFKTEEE 404

Db 373 YIQLGQKEGAKLDCGGERGFFIKPTVFGGVQDDMRIAKEEIFGPPVQPLPKFKKIEE 432

QY 405 AIEIANDTTYGLASVIYTKSLNRGLRVSSALETGGVSINFPFIPETQTPEFGMKQSGSGR 464

Db 433 VVERANNTRYGLAAAVFTRDLKAMYFTQALQAGTVWVNTYNIIVTCHTPEFGPKESNGR 492

QY 465 ELGEEGLKAYLEPKTINIHV 484

Db 493 ELGEDGLKAYTEVKTIVKV 512

RESULT 2

S43114

aldehyde dehydrogenase (NAD+) (EC 1.2.1.3) - fungus (Cladosporium herbarum)

C;Species: Cladosporium herbarum

C;Date: 19-Mar-1997 #sequence_revision 06-Jun-1997 #text_change 21-Jan-2000

C;Accession: S43114

R;Achatz, G.; Oberkofler, H.; Simon, B.; Lechenauer, E.; Unger, A.; Kandler, D.; Prillinger submitted to the EMBL Data Library, March 1994

A;Description: Molecular characterization of allergens of Cladosporium herbarum and Alternaria alternata

A;Reference number: S43108

A;Accession: S43114

A;Molecule type: mRNA

A;Residues: 1-496 <ACH>

A;Cross-references: EMBL:X78228; NID:g467624; PIDN:CAA55072.1; PID:g467625

C;Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

C;Keywords: NAD; oxidoreductase

F;56-317/Domain: aldehyde dehydrogenase homology <ALD>

Query Match 34.0%; Score 866; DB 2; Length 496;

Best Local Similarity 42.8%; Pred. No. 3.7e-60;

Matches 211; Conservative 67; Mismatches 183; Indels 32; Gaps 12;

QY 7 EYKSELFINNEFVSSKGSERLTITNPWDESTVATDVHVAANAADVDSAVAASVQAVKKGPW 66

Db 15 EQPTGLFINNEFVKGQEGKTFDVINPSDES-VITQVHEATEKDVDIAVAARQAF-EGSW 72

QY 67 KKFTGAQRAACMLKFADLAEKNAEKLARLESPLTGRPVSMITHFDIPNMVSV-----FRY 121

Db 73 RLETPENRGKLLNNLANLFKKNLTLAAVESLDNGKATSM-----ARVTSACAGCLRY 126

QY 122 YAGWADKIAGKTFPEDNGKPNWRY---EPMGVCAGIASWNATFLYVGVKWIAPALAAAGCSF 178

Db 127 YGGWADKITGKVI--DTTPDTFNKYKKEPIGVCRSDHSELPPLLMWAWKIGPAIACGNTV 184

QY 179 IFKASEKSPLGLVGLAPLFAEAGFPFGVQFLTG-ARVTGEALASHMDIAKISFTRSVGG 237

Db 185 VLKTAEOPTPLGLVAASLVKEAGFPFCVINVISCGFKVAGAAALSSHMDVDKVAFTGSTVV 244

QY 238 GRAVKQATLKSNMKRVTLLELGEK-PTIVFNEAPLERQ-SGESAKDFSKFGQIWPVPPSCLL 295

Db 245 GRTILKAAASNLKKVTLLELGGKSPNIVFEDADIDNAISWVNFEGIFFNHGQCCAGSRVY 304

QY 296 VQ---WGNLAEKPHGVRHGSFGGRCORWLQONPLEPKRTHGPFVDKSYDRVLGNIDVGKD 352

Db 305 VQESIYDKFVQKFK-----ERAQKNVVGDPFAADTFQGPQVSKVQFDRIMEYIQAGKD 357

QY 353 T-AQLLTGVRGKDGKGFATIEPTIFVNPKPCKSIWFEEIFGPPVLSIKTFKTEEEAIEIAND 411

Db 358 AGATVETGSRKRGDKGYFIEPTIPSNVNTEDMKIVKEEIFGPPVCSIAKFKTKEDAIKLQNA 417

QY 412 TTYGLASVIYTKSLNRGLRVSSALETGGVSINFPFIPETQTPFGCMKQSGSGRELGEGL 471

Db 418 STYGLAAAVHTKNLNTAIEVSNALKAGTVVWNTYNTLHHQMPFGGKESGIGRELGEDAL 477

QY 472 KAYLEPKTINIHV 484

Db 478 ANYTQTKTVSIRL 490

RESULT 3

A29055

aldehyde dehydrogenase (NAD+) (EC 1.2.1.3) - Emericella nidulans

C;Species: Emericella nidulans, Aspergillus nidulans

C;Date: 30-Jun-1988 #sequence_revision 31-Jan-1997 #text_change 11-Jun-1999

C;Accession: A29055

R;Pickett, M.; Gwynne, D.I.; Buxton, F.P.; Elliott, R.; Davies, R.W.; Lockington, R. Gene 51, 217-226, 1987

A;Title: Cloning and characterization of the aldA gene of Aspergillus nidulans.

A;Reference number: A29055; MUID:87248080

A;Accession: A29055

A;Molecule type: DNA

A;Residues: 1-497 <PIC>

A;Cross-references: GB:M16197; NID:g168010; PIDN:AAA33293.1; PID:g168011

A;Experimental source: strain Glasgow FGSC4

C;Genetics:

A;Gene: aldA

A;Map position: VIII

A;Introns: 21/2; 422/1

C;Function:

A;Description: catalyzes the oxidation of an aldehyde to an acid using NAD+ and water.

A;Pathway: ethanol catabolism

A;Note: enzymes with this activity are involved in diverse metabolic pathways in various families: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

C;Superfamily: aldehyde dehydrogenase (NAD+); oxidoreductase

C;Keywords: alcohol metabolism; NAD; oxidoreductase

F;57-318/Domain: aldehyde dehydrogenase homology <ALDD>

F;264,298/Active site: Glu, Cys #status predicted

Query Match 33.9%; Score 864; DB 1; Length 497;

Best Local Similarity 42.3%; Pred. No. 5.3e-60;

Matches 203; Conservative 79; Mismatches 176; Indels 22; Gaps 11;

QY 12 LFINNEFVSSKGSERLTITNPWDESTVATDVHVAANAADVDSAVAASVQAVKKGPWKFTG 71

Db 21 LFINNEFVKGVEGKTFQVINPSNEK-VITSYHEATEKDVDVAVAAA-RAAFEGPWRQVTP 78

QY 72 AORAAACMLKFADLAEKNAEKLARLESPLTGRPVSMITHFDIPNMVSVFERYYAGWADKIAG 131

Db 79 SERGILINKLADLMERDIDTLAAIESLDNGKAFTM-AKVDLANSICGLRYVAGWADKIRG 137

QY 132 KTFPEDNGKPNW-RYEPMGVCAGIASWNATFLYVGVKWIAPALAAAGCSFIFKASEKSPLGV 190

Db 138 QTIDTNPETLTYYTRHEPVGVCQGIIPWNPFLMWSWKIGPAVAAGNTVVLKTAQQTPLSA 197

QY 191 LGLAPLFAEAGFPFGVQFLTG-ARVTGEALASHMDIAKISFTRSVGGGGRVAKQATLKS 249

Db 198 LYAAKLIKEAPFPAGVINVISGFGRTAGAAISSHMDIDKVAFTGSTLVGPTILQAAAKSN 257

QY 250 MKRVTLLELGEK-PTIVFNEAPLERQ-SGESAKDFSKFGQIWPVPPSCLLVOMGNLAEKF-- 305

Db 258 LKKVTLLEGGKSPNIVFDDADIDNAISWVNFEGIFFNHGQCCAGSRILVQEG-IYDKFVA 316

QY 306 ---HGVRHGSFGGRCORWLQONPLEPKRTHGPFVDKSYDRVLGNIDVGKDT-AQLLTG 361

Db 317 RPKERAQKNKVG-----NPFQDFTQGPQVSQLQFDRIMEYINHGKKAGATVATGGD 368

QY 362 RKGDGGAIEPTIFVNPKPCKSIWFEEIFGPPVLSIKTFKTEEEAIEIANDTTYGLASVIY 421

Db 369 RHNEGYFIQPTVFTDVTSDMKIAQEEIFGPPVVTIQKFDVAEAIKIGNSTDYGLAAAVH 428

QY 422 TKSLSNRGLRVSSALETGGVSINFPFIPETQTPFGCMKQSGSGRELGEGLKAYLEPKTIN 481

Db 429 TKNVNTAIRVSNALKAGTVWVNNYNNMISYQAPFGGFKQSGLRELGSYALENTYQIKTVH 488

RESULT 4

aldehyde dehydrogenase (NAD+) (EC 1.2.1.3) 1 precursor, cytosolic - mouse

N;Alternate names: aldehyde dehydrogenase II

C;Species: Mus musculus (house mouse)

C;Date: 23-Nov-1991 #sequence_revision 31-Jan-1997 #text_change 11-Jun-1999

C;Accession: JQ1004; A33248; A60126

R;Rongnoparut, P.; Weaver, S.

Gene 101, 261-265, 1991

A;Title: Isolation and characterization of a cytosolic aldehyde dehydrogenase-encoding c

A;Reference number: JQ1004; MUID:91276281

A;Accession: JQ1004

A;Molecule type: mRNA

A;Residues: 1-501 <RON>

A;Cross-references: GB:M74570; NID:g191803; PIDN:AAA37202.1; PID:g191804

A;Experimental source: liver

A;Accession: A33248

A;Molecule type: DNA

A;Residues: 1-22 <RO2>

A;Cross-references: GB:M74571; NID:g191805; PIDN:AAA37203.1; PID:g191806

R;McCaffery, P.; Tempst, P.; Lara, G.; Draeger, U.C.

Development 112, 693-702, 1991

A;Title: Aldehyde dehydrogenase is a positional marker in the retina.

A;Reference number: A60126; MUID:92037178

A;Accession: A60126

A;Molecule type: protein

A;Residues: 23-49,'XX',52;'Y',141-150,'X',152-156;211,'XX',214-218,'X',220-230;309-320;3

A;Experimental source: embryonic mouse retina

C;Genetics:

A;Gene: Aldh1; Ahd-2

A;Map position: 19

C;Complex: homotetramer

C;Function:

A;Description: catalyzes the oxidation of an aldehyde to an acid using NAD+ and water

A;Pathway: oxidation of retinaldehyde to retinoic acid

A;Note: enzymes with this activity are involved in diverse metabolic pathways in various

C;Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

C;Keywords: acetylated amino end; alcohol metabolism; cytosol; homotetramer; liver; NAD;

F;2-501/Product: aldehyde dehydrogenase (NAD+) 1 #status predicted <NAT>

F;2-323/Domain: aldehyde dehydrogenase homology <ALDD>

F;2/Modified site: acetylated amino end (Ser) (in mature form) #status predicted

F;269,303/Active site: Glu, Cys #status predicted

F;456/Binding site: NAD (Cys) #status predicted

Query Match 33.6%; Score 856.5; DB 1; Length 501;

Best Local Similarity 40.8%; Pred. No. 2.1e-59;

Matches 197; Conservative 87; Mismatches 186; Indels 13; Gaps 8;

QY 10 SELFINNEFVSSKGSERLTLTNPWDESTVATDVHVANAADVDSAAVASQAVKKG-PWKK 68

Db 21 TKIFINNEWHNSVSGKKFPVLNPATEEVIC-HVEEGDKADVDRKAVKAARQAFQIGSPWRT 79

QY 69 FTGAQRAACMLKFADLAEKNAEKLARLESPTGRPVSMITHFDIPNMVSVFRYYAGWADK 128

Db 80 MDASERGLLNKLADLMERDRLLATMEALNGKVFANAYLSDLGGCIKALKYCAGWADK 139

QY 129 IAGKTFPEDNGKPNW-RYEPMGVCAGIASNNATFLYVGWKIAPALAAAGCSFIFKASEKSP 187

Db 140 IHGQTIPSDGDIITYTRREPVGCGQIIPNFPMLFIWKIGPALSCGNTVVVKPAEQTP 199

QY 188 LGVLGLAPLFAEAGFPVGVQFLTGARVT-GEALASHMDIAKISFTRSVGGGRAVKQATL 246

Db 200 LTAHLASLIKEAGFPVGVNIVPGYPTAGAAISSHMDVDKVAFTGSTQVGKLIKEAAG 259

QY 247 KSNMKRVTTLELGEK-PTIVFNEAPLERQSGESAKDFSKEGQIWPVPPSCLLVQWGNLAEK- 304

Db 260 KSNLKRVTLELGGKSPCIFVADADL-----DIAVEFAHHGVFVHQGCCVAAASRIFVEES 314

QY 305 -FHGVRHGSFGGCGQRLGNQNLPEPKRTHGPFVDKSOYDRVLGNIDVG-KDTAQLLTGVGR 362

Db 315 VYDEFVKRSVERAKKYVLGNPLTPGINOGPOIDKEHQDKILDILIESGKKEGAKLECGGR 374

QY 363 KGDKGFALEPTIFVNPKPGRSKIWFEEIFGPVLSIKTFKTEEEAIEIANDTTYGLASVIYT 422

Db 375 WGNKGFFVQPTVFSNVTDEMRIAEEIFGPVQQIMKFKSVDDVIKRANNTTYGLAAGLFT 434

QY 423 KSLNRGLRVSSALETGGVSINFPFIPETQTPFGGKMKOSGSGRELGEGLKAYLEPKTINI 482

Db 435 KDLDKAITVSSALQAGVWVWNCYIMLSAQCPFGGFKMSNGRELGEHGLYEYTELKTVM 494

QY 483 HVN 485

Db 495 KIS 497

RESULT 5

S02302

aldehyde dehydrogenase (NAD+) (EC 1.2.1.3) 1, cytosolic - horse

C;Species: Equus caballus (domestic horse)

C;Date: 01-Dec-1989 #sequence_revision 31-Jan-1997 #text_change 31-Jan-1997

C;Accession: S02302

R;von Bahr-Lindstroem, H.; Hempel, J.; Joernvall, H.

Eur. J. Biochem. 141, 37-42, 1984

A;Title: The cytoplasmic isoenzyme of horse liver aldehyde dehydrogenase. Relationshi

A;Reference number: S02302; MUID:84208025

A;Accession: S02302

A;Molecule type: protein

A;Residues: 1-500 <VON>

A;Note: residues in the regions 1-16 and 443-475 were positioned by homology with the

C;Complex: homotetramer

C;Function:

A;Description: catalyzes the oxidation of an aldehyde to an acid using NAD+ and water

A;Note: enzymes with this activity are involved in diverse metabolic pathways in vari

C;Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

C;Keywords: acetylated amino end; alcohol metabolism; cytosol; homotetramer; liver; N

F;58-322/Domain: aldehyde dehydrogenase homology <ALDD>

F;1/Modified site: blocked amino end (Ser) (probably acetylated) #status experimental

F;268,302/Active site: Glu, Cys #status predicted

F;455/Binding site: NAD (Cys) #status predicted

Query Match 33.6%; Score 855.5; DB 1; Length 500;

Best Local Similarity 41.9%; Pred. No. 2.5e-59;

Matches 202; Conservative 87; Mismatches 182; Indels 11; Gaps 10;

QY 10 SELFINNEFVSSKGSERLTLTNPWDESTVATDVHVANAADVDSAAVASQAVKKG-PWKK 68

Db 20 TKIFINNEWHDSVSGKKFPVFNPAEKLIC-EVEEGDKEDVNAKAVAAARQAFQIGSPWRT 78

QY 69 FTGAQRAACMLKFADLAEKNAEKLARLESPTGRPVSMITHFDIPNMVSVFRYYAGWADK 128

Db 79 MDASERGLLYKLADLVERDRLLATMESMNGGKLFSNAYLMDLGGCLKTLRYCAGWADK 138

QY 129 IAGKTFPED-NGKPNWRYEPMGVCGAGIASNNATFLYVGWKIAPALAAAGCSFIFKASEKSP 187

Db 139 IQGRTIPSDGNFETTRHEPVGVCGQILPNWFPLLMFLWKIAPALSCGNTVVVKPAEQTP 198

QY 188 LGVLGLAPLFAEAGFPVGVQFLTGARVT-GEALASHMDIAKISFTRSVGGGRAVKQATL 246

Db 199 LSALHVAATLIKEAGFPVGVNIVPGYPTAGAAISSHMDIDKVAFTGSTEVGKLIKEAAG 258

QY 247 KSNMKRVTTLELGEK-PTIVFNEAPLERQSGESAKD-FSKFGQIWPVPPSCLLVQWGNLAEK 304

Db 259 KSNLKRVTLELGGKSPFIVFADADLLETALEVT HQALFYHOGCCVAAASRLFVE-ESIYDE 317

QY 305 FHGVRHGSFGGCGQRLGNQNLPEPKRTHGPFVDKSOYDRVLGNIDVG-KDTAQLLTGVGRK 363

Db 318 F--VRR-SVERAKKYVLGNPLTPGVSQGPQIDKEQYDKILDILIESGKKEGAKLECGGCPW 374

QY 364 GDKGFAIEPTIFVNPKPGRSKIWFEEIFGPVLSIKTFKTEEEAIEIANDTTYGLASVIYTK 423

Db 375 GNKGYFIQPTVFSNVSDVMRIAKEEIFGPVQIMKFKSLDDVIKRANNTTYGLFAGSFTK 434

N:Alternate names: aldehyde dehydrogenase E1
C:Species: Homo sapiens (man)
C:Date: 28-May-1986 #sequence_revision 31-Dec-1993 #text_change 08-Dec-2000
C:Accession: A33371; I51846; I39431; JC5554; A00363; A61010; A26817
R:Hsu, L.C.; Chang, W.C.; Yoshida, A.
Genomics 5, 857-865, 1989
A:Title: Genomic structure of the human cytosolic aldehyde dehydrogenase gene.
A:Reference number: A33371; MUID:90077427
A:Accession: A33371
A:Molecule type: DNA
A:Residues: 1-501 <HSU>
A:Cross-references: GB:M31994; NID:gl78370; PIDN:AAA51692.1; PID:gl78372
R:Yoshida, A.; Hsu, L.C.; Yanagawa, Y.
Adv. Exp. Med. Biol. 328, 37-44, 1993
A:Title: Biological role of human cytosolic aldehyde dehydrogenase 1: hormonal response.
A:Reference number: I51846; MUID:93263033
A:Accession: I51846
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-6 <YOS1>
A:Cross-references: GB:S61235; NID:g300399; PIDN:AAD13925.1; PID:g4261625
R:Yoshida, A.; Ikawa, M.; Hsu, L.C.; Tani, K.
Alcohol 2, 103-106, 1985
A:Title: Molecular abnormality and cDNA cloning of human aldehyde dehydrogenases.
A:Reference number: I39431; MUID:85252089
A:Accession: I39431
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: '1', 163-501 <YOS2>
A:Cross-references: GB:M26761; NID:gl78393; PIDN:AAA35518.1; PID:gl78394
R:Kathmann, E.C.; Lipsky, J.J.
Biochem. Biophys. Res. Commun. 236, 527-531, 1997
A:Title: Cloning of a cDNA encoding a constitutively expressed rat liver cytosolic aldehyde dehydrogenase.
A:Reference number: JC5553; MUID:97382470
A:Accession: JC5554
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 1-120, 'S', 122-501 <KAT>
A:Cross-references: DBJ:AF003341; NID:g2183298; PIDN:AAC51652.1; PID:g2183299
A:Experimental source: liver
R:Hempel, J.; von Bahr-Lindstrom, H.; Jornvall, H.
Eur. J. Biochem. 141, 21-35, 1984
A:Title: Aldehyde dehydrogenase from human liver. Primary structure of the cytoplasmic isozyme.
A:Reference number: A00363; MUID:84208019
A:Accession: A00363
A:Molecule type: protein
A:Residues: 2-501 <HEM>
R:Agarwal, D.P.; Cohn, P.; Goedde, H.W.; Hempel, J.
Enzyme 42, 47-52, 1989
A:Title: Aldehyde dehydrogenase from human erythrocytes: structural relationship to the aldehyde dehydrogenase (NAD(P)+) (EC 1.2.1.5) - rat
A:Reference number: A61010; MUID:89377753
A:Accession: A61010
A:Molecule type: protein
A:Residues: 23-27; 79-85; 101-107; 114-128; 132-142; 144-154; 309-319; 421-434; 477-483 <AGA>
A:Experimental source: erythrocytes
R:Abriola, D.P.; Fields, R.; Stein, S.; MacKerell Jr., A.D.; Pietruszko, R.
Biochemistry 26, 5679-5684, 1987
A:Title: Active site of human liver aldehyde dehydrogenase.
A:Reference number: A26817; MUID:88050861
A:Accession: A26817
A:Molecule type: protein
A:Residues: 266-273 <ABR>
A:Note: the active site Glu was identified by suicide inhibition with bromoacetophenone
C:Genetics:
A:Gene: GDB:ALDH1
A:Cross-references: GDB:119667; OMIM:100640
A:Map position: 9q21-9q21
C:Complex: homotetramer
C:Function:
A:Description: catalyzes oxidation of an aldehyde to an acid using NAD+ and water
A:Note: enzymes with this activity are involved in diverse metabolic pathways in various tissues
C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
C:Keywords: acetylated amino end; alcohol metabolism; cytosol; homotetramer; liver; NAD;

F:2-501/Product: aldehyde dehydrogenase (NAD+) 1 #status experimental <MAT>
F:59-323/Domain: aldehyde dehydrogenase homology <ALDD>
F:2/Modified site: acetylated amino end (Ser) (in mature form) #status experimental
F:269,303/Active site: Glu, Cys #status experimental
F:456/Binding site: NAD (Cys) #status experimental
Query Match 33.1%; Score 843.5; DB 1; Length 501;
Best Local Similarity 39.8%; Pred. No. 2.2e-58;
Matches 192; Conservative 90; Mismatches 188; Indels 13; Gaps 8;
QY 10 SELFNNFVSSKGSERLTLPNPWDESTVATDVHVANAADVDSAVAASVQAVKKG-PWKK 68
Db 21 TKIFINNWHDSVSGKKFPVPNPATEEELC-QVEEGDKEDVDKAVKAARQAFQIGSPWRT 79
QY 69 FTGAQRAACMLKFADLAEKNAEKLESLPTGRPVSMITHFDIPNMVSVFRYYAGWADK 128
Db 80 MDASERGRLLYKLADLIERDRLLLATMESMNGGKLYSNAYLNDLAGCIKTLRYCAGWADK 139
QY 129 IAGKTFPED-NGKPNWRYEPNGVCAGIASWNATFLYVGVWKIAPALAAAGCSFIFKASEKSP 187
Db 140 IQGRTIPIDGNFFTYTRHEPIGVCGQIIPWNFPLVMLIWKIGPALSCGNTVTVVVKPAEQTP 199
QY 188 LGVLGLAPLFAEAGFPPGVVQFLTGARVT-GEALASHMDIAKISFTRSVGGGGRVAKQATL 246
Db 200 LTALHVASLIKEAGFPPGVVNIIVPGYGTAGAAISSHMDIDKVAFTGSTEVGKLIKEAAG 259
QY 247 KSNMKRVTLLELGEK-PTIVFNEAPLERQSGESAKDFSKFGQIWWPPSCLLVQWGNLAEK- 304
Db 260 KSNLKRVTLELGGKSPCIVLADADL-----DNAVEFAHHGVFVHQGCCIAASRIFVEES 314
QY 305 -FHGVRHGSFGGCGQWLQGNPLEPKRTHGPFYVKSQYDRVLGNIDVG-KDTAQLLTGVGR 362
Db 315 IYDEFVRRSVERAKKYILGNPLTPGVQTGPQIDKEQYDKILDLESKKKEGAKLECGGGP 374
QY 363 KGDKGEAIEPTIFVNPKPCKSKIMFEEIFGPVLSIKTFKTEEEAIEIANDTTYGLASVIYT 422
Db 375 WGNKGVFVQPTVPSNVTDENRIAKEEIEFGPVQIQIMKFKSLDDVIKRANNTFYGLSAGVFT 434
QY 423 KSLNRGLRVSSALETGGVSINFPFIPETQTPFGMKQSGSGRELGEGLKAYLEPKTINI 482
Db 435 KDIDKAITISSALQAGTVVWVNCYGVVSAQCFFGFGKMSGNGRELGEYGFHEYTEVKTVTV 494
QY 483 HVN 485
Db 495 KIS 497
RESULT 9
JC4524
aldehyde dehydrogenase (NAD(P)+) (EC 1.2.1.5) - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 08-Feb-1996 #sequence_revision 19-Apr-1996 #text_change 24-Nov-1999
C:Accession: JC4524; PC4119
R:Bhat, P.V.; Labrecque, J.; Boutin, J.M.; Lacroix, A.; Yoshida, A.
Gene 166, 303-306, 1995
A:Title: Cloning of a cDNA encoding rat aldehyde dehydrogenase with high activity for
A:Reference number: JC4524; MUID:96125208
A:Accession: JC4524
A:Molecule type: mRNA
A:Residues: 1-501 <BHA>
A:Cross-references: GB:L42009; NID:gl256958; PIDN:AAA96657.1; PID:g974168
A:Accession: PC4119
A:Molecule type: protein
A:Residues: 2-19; 80-84; 86; 88-91; 96-99; 101-108; 110-121; 205-209; 213-215; 218-219; 222; 22
A:Experimental source: Kidney
C:Genetics:
A:Gene: aldh
C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
C:Keywords: acetylated amino end; kidney; NAD; oxidoreductase
F:2-501/Product: aldehyde dehydrogenase #status predicted <MAT>
F:59-323/Domain: aldehyde dehydrogenase homology <ALDD>
F:2/Modified site: acetylated amino end (Ser) (in mature form) #status predicted

Query Match 33.1%; Score 843.5; DB 2; Length 501; Best Local Similarity 40.4%; Pred. No. 2.2e-58; Matches 195; Conservative 86; Mismatches 189; Indels 13; Gaps 8;									
QY	10	SELF	FINNEFVSSKGSERLTLTNPWDESTVATDVHVANRADVDSAVAASVQAVKKG-PWKK	68					
			:::	1	:	:	:	:	:
Db	21	TKIF	INNEHDSVSGKKFPVLNPATEEVIC-HVEEGKADVDKAVKAARQAFQIGSPWRT	79					
			:::	1	:	:	:	:	:
QY	69	FTGA	QRAACMLKFAADLAEKNAEKLARLESPTGRPVSMITHFDIPNMVSFRYYAGWADK	128					
			:::	1	:	:	:	:	:
Db	80	MDAS	ERGRLLKLADLMERDCLLLATIEAINGGKVAFANAYLSDLGGSIKALKYCAGWADK	139					
			:::	1	:	:	:	:	:
QY	129	IAGK	TFPEDNGKPNW-RYEPMGVCAGIASWNATFLYVGWKIAPALAAAGCSFIFKASEKSP	187					
				1	:	:	:	:	:
Db	140	IHGQ	TIPSDGDIETTRREPIGVCGQIIPWEFPLLMFIWKIGPALSCGNTVVVKPAEQTP	199					
				1	:	:	:	:	:
QY	188	LGVL	GLAPLEAEAGFPVGVQFLTGARVT-GEALASHMDIAKISFTRSVGGGRAVKQATL	246					
				1	:	:	:	:	:
Db	200	LTAL	HMASLIXEAGFPVGVNIVPGYGTAGAAISSHMDVDKVAFSTQVQKLIKEAAG	259					
				1	:	:	:	:	:
QY	247	KSNM	KRVTLLEGEK-PTIVNEAPLERQSGESAKDFSKEGQIWPVPPSCLLVQWGNLAEK-	304					
				1	:	:	:	:	:
Db	260	KSNL	KRVTLLEGGKSPCIVFADADL-----DIAVEFAHHGVVFYHOGQCCVAAASRIFVEES	314					
				1	:	:	:	:	:
QY	305	-FHG	VRHSGFCGQRLWQGNPLEPKRTHGPFVDKSOYDRVLGNIDVG-KDTAQLLTGVGR	362					
				1	:	:	:	:	:
Db	315	VYDE	FVRKSVERRAKKYVLGNPLTQGINQGPQIDKEQDKILDLESKKEGAKLECGGGR	374					
				1	:	:	:	:	:
QY	363	KGDK	GFAIEPTIFVNPKPGSKIWFEEIFGPVLSIKTFKTEEEAIEIANDTTYGLASVIYT	422					
				1	:	:	:	:	:
Db	375	WGNK	GFVQPIVFSNVTDEMRIAKEEIFGPVQOIMKFKSIDDIKVRANNTTYGLAAGVFT	434					
				1	:	:	:	:	:
QY	423	KSLN	RGLRVSSALETGGVSINFPFIPETQTPFGGMKQSGSGRELGEGLKAYLEPKTINI	482					
				1	:	:	:	:	:
Db	435	KOLD	RAITVSSALQAGVVWVNCYMILSAQCFFGGFKMSGNGRELGEHGLYEYTELKTVM	494					
				1	:	:	:	:	:
QY	483	HVN	485						
			:						
Db	495	KIS	497						
			:						
RESULT 10									
DEHUE2									
aldehyde dehydrogenase (NAD+) (EC 1.2.1.3) 2 precursor, mitochondrial [validated] - huma									
N:Alternate names: aldehyde dehydrogenase E2; aldehyde dehydrogenase I									
C:Species: Homo sapiens (man)									
C:Date: 31-Mar-1988 #sequence_revision 17-Nov-1995 #text_change 08-Dec-2000									
A:Accession: A29975; S00804; A23503; A27509; A26743; I39432; I39433; I39430									
R:Hsu, L.C.; Bendel, R.E.; Yoshida, A.									
Genomics 2, 57-65, 1988									
A:Title: Genomic structure of the human mitochondrial aldehyde dehydrogenase gene.									
A:Reference number: A29975; MUID:88256152									
A:Accession: A29975									
A:Molecule type: DNA									
A:Residues: 1-517 <HSU1>									
A:Cross-references: GB:M26760; NID:g178397; PIDN:AAA51694.1; PID:g178398									
R:Braun, T.; Bober, E.; Singh, S.; Agarwal, D.P.; Goedde, H.W.									
FEBS Lett. 233, 440, 1988									
A:Reference number: S00804									
A:Accession: S00804									
A:Molecule type: mRNA									
A:Residues: 1-6, 'ARA', 10, 'P', 13-49 <BRA>									
A:Cross-references: EMBL:X05409; NID:g28605									
A:Note: correction to A26743									
R:Hempel, J.; Kaiser, R.; Joernvall, H.									
Eur. J. Biochem. 153, 13-28, 1985									
A:Title: Mitochondrial aldehyde dehydrogenase from human liver. Primary structure, diffe									
A:Reference number: A23503; MUID:86055846									
A:Accession: A23503									
A:Molecule type: protein									
A:Residues: 'A', 19-517 <HEM>									
A:Note: the sequence shown is presumably that of the mature protein; however, the n									
R:Braun, T.; Bober, E.; Singh, S.; Agarwal, D.P.; Goedde, H.W.									
Nucleic Acids Res. 15, 3179, 1987									
A:Title: Isolation and sequence analysis of a full length cDNA clone coding for hum									
A:Reference number: A27509; MUID:87174836									
A:Accession: A27509									
A:Molecule type: mRNA									
A:Residues: 1-6, 'AWPAA', 10, 'P', 12, 'VS', 15, 'RHPGR', 21, 27-79, 'REGRPG', 86-336, 'V', 338									
A:Cross-references: EMBL:Y00109; NID:g28607; PIDN:CAA68290.1; PID:g28608									
R:Braun, T.; Bober, E.; Singh, S.; Agarwal, D.P.; Goedde, H.W.									
FEBS Lett. 215, 233-236, 1987									
A:Title: Evidence for a signal peptide at the amino-terminal end of human mitochond									
A:Reference number: A26743; MUID:87219091									
A:Accession: A26743									
A:Molecule type: mRNA									
A:Residues: 1-6, 'AWPAA', 10, 'P', 12, 'VS', 15, 'RHPGR', 21, 27-79, 'REGRPG', 86-336, 'V', 338									
A:Cross-references: EMBL:X05409; NID:g28605									
A:Note: this sequence is revised in reference S00804									
R:Hsu, L.C.; Tani, K.; Fujiyoshi, T.; Kurachi, K.; Yoshida, A.									
Proc. Natl. Acad. Sci. U.S.A. 82, 3771-3775, 1985									
A:Title: Cloning of cDNAs for human aldehyde dehydrogenases 1 and 2.									
A:Reference number: I39432; MUID:85216574									
A:Accession: I39432									
A>Status: preliminary									
A:Molecule type: mRNA									
A:Residues: 119-336, 'V', 338-517 <HSU2>									
A:Cross-references: GB:K03001; NID:g178395; PIDN:AAB59500.1; PID:g178396									
A:Note: thirty-three tryptic peptides were also sequenced									
R:Yoshida, A.; Ikawa, M.; Hsu, L.C.; Tani, K.									
Alcohol 2, 103-106, 1985									
A:Title: Molecular abnormality and cDNA cloning of human aldehyde dehydrogenases.									
A:Reference number: I39431; MUID:85252089									
A:Accession: I39433									
A>Status: preliminary; translated from GB/EMBL/DBJ									
A:Molecule type: mRNA									
A:Residues: 119-336, 'V', 338-517 <YOS>									
A:Cross-references: GB:M26760; NID:g178397; PIDN:AAA51694.1; PID:g178398									
R:Agarwal, D.P.; Goedde, H.W.									
Isozymes Curr. Top. Biol. Med. Res. 16, 21-48, 1987									
A:Title: Human aldehyde dehydrogenase isozymes and alcohol sensitivity.									
A:Reference number: I39430; MUID:87279033									
A:Accession: I39430									
A>Status: preliminary; translated from GB/EMBL/DBJ									
A:Molecule type: mRNA									
A:Residues: 214-215, 'S', 217, 'R', 219-246, 'P', 248-336, 'V', 338-425, 'EVQDHRGGCWESQQFHVRA									
A:Cross-references: GB:M54931; NID:g178391; PIDN:AAA62825.1; PID:g178392									
A:Note: the sequence is misidentified as aldehyde dehydrogenase I									
C:Genetics:									
A:Gene: GDB:ALDH2									
A:Cross-references: GDB:119668; OMIM:100650									
A:Map position: 12q24.2-12q24.2									
A:Introns: 38/3; 73/3; 120/3; 147/2; 184/3; 227/3; 265/3; 300/1; 361/2; 416/3; 469/2									
C:Complex: homotetramer									
C:Function:									
A:Description: catalyzes the oxidation of an aldehyde to an acid using NAD+ and wate									
A:Pathway: ethanol catabolism									
A:Note: enzymes with this activity are involved in diverse metabolic pathways in var									
C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase.homology									
C:Keywords: alcohol metabolism; homotetramer; liver; mitochondrion; NAD; oxidoreduct									
F:1-17/Domain: transit peptide (mitochondrion) #status predicted <SIG>									
F:18-517/Product: aldehyde dehydrogenase (NAD+) 2 #status experimental <MAT>									
F:75-339/Domain: aldehyde dehydrogenase homology <ALDD>									
F:211-291/Domain: NAD binding #status predicted <NAD>									
F:285/Active site: Glu #status predicted									
F:319/Active site: Cys #status experimental									
F:472/Binding site: NAD (Cys) #status predicted									
Query Match 33.0%; Score 840.5; DB 1; Length 517; Best Local Similarity 41.0%; Pred. No. 3.9e-58; Matches 200; Conservative 86; Mismatches 187; Indels 15; Gaps 12;									
QY	5	PDEYK	SELFINNFEVSSKGSERLTLTNPWDESTVATDVHVANRADVDSAVAASVQAVKKG	64					

Db 32 PEVFCNQIFINNEWHDAVSRKTEPTVNP-STGEVICQVAEGDKEDVDKAVKAARAAQQLG 90
QY 65 -PWKKFTGAQRAACMLKFADLAEKNAEKLARLESPTGRPVSMITHEDIPNMVSFRYYA 123
Db 91 SPWRRMDASHRGRLLNRLADLIERDRTYLALETLDNGKPVVISYLVLDMLVKCLRYA 150
QY 124 GWADKIAGKTFPEDNGKPNW-RYEPMGVCAGIASWNATFLYVGWKIAPALAAAGCSFIEKA 182
Db 151 GWADKYHGKTIPIDGDFSYTRHEPVGCGQIIPWNFPLLMQAWKLGALATGNVVMKV 210
QY 183 SEKSPLGLVLGLAFABAGFPVGVVQFLTGARVT-GEALASHMDIAKISFTRSVGGGRAV 241
Db 211 AEQTPLTALYVANLIKEAGFPVGVVNIVPFGPTAGAAIASHEDVDKVAFTGSTEIGRVI 270
QY 242 KOATLKSNMKRVTLLELGEK-PTIVFNEAPLERQSGESAKDFSKF---GQIWVPPSCLLVQ 297
Db 271 QVAAGSSNLKRVTLLELGGKSPNIIMSDADMD-WAVEQA-HFALFFNQGCCACGSRFTVQ 328
QY 298 WGNLAEKFGVHRHSGSGGCQRLWGNLQNPLEPKRTHGPFVDKSOYDRVLGNIDVGK-DTAQL 356
Db 329 -EDIYDEF--VERSARAKSRVVG-NPFDSKTEQGPQVDETQFKILGYINTGKQEGAKL 384
QY 357 LTGVGRKDGKGFALIEPTIFVNPKPQSGKIWFEEIFGPVLSIKTFKTEEEAIEIANDTTYGL 416
Db 385 LCGGGIAADRGYFIQPTVFGDVQDGMTIAKEEIFGPVMQILKFKTIEEVVGRANNSTYGL 444
QY 417 ASVIYTKSLNRGLRVSSALETGGVSINFPFIPETQTPFGGMKQSGSGRELGEGLKAYLE 476
Db 445 AAAVFTKDLKANYLSQALQAGTVWVNCYDVFGAQSPFGGYKMSGSGRELGEYGLQAYTE 504
QY 477 PKTINIHV 484
Db 505 VKTWTVKV 512

RESULT 11
S14629
aldehyde dehydrogenase (NAD+) (EC 1.2.1.3) 1, cytosolic - chicken
C:Species: Gallus gallus (chicken)
C:Date: 21-Nov-1993 #sequence_revision 31-Jan-1997 #text_change 11-Jun-1999
C:Accession: S14629
R:Godbout, R.
Submitted to the EMBL Data Library, April 1991
A:Description: High levels of aldehyde dehydrogenase transcripts in the undifferentiated
A:Reference number: S14629
A:Accession: S14629
A:Molecule type: mRNA
A:Residues: 1-509 <GOD>
A:Cross-references: EMBL:X58869; NID:963032; PIDN:CAA41679.1; PID:963033
A:Experimental source: retina
C:Complex: homotetramer
C:Function:
A:Description: catalyzes the oxidation of an aldehyde to an acid using NAD+ and water
A:Pathway: ethanol catabolism
A:Note: enzymes with this activity are involved in diverse metabolic pathways in various
C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
C:Keywords: alcohol metabolism; cytosol; homotetramer; liver; NAD; oxidoreductase
F:67-331/Domain: aldehyde dehydrogenase homology <ALDD>
F:277,311/Active site: Glu, Cys #status predicted
F:464/Binding site: NAD (Cys) #status predicted

Query Match 32.9%; Score 838.5; DB 1; Length 509;
Best Local Similarity 39.8%; pred. No. 5.5e-58;
Matches 192; Conservative 87; Mismatches 190; Indels 13; Gaps 8;

QY 10 SELFFINNEFVSSKGSRLTLTNPWDESTVATDVHANAADVSAVAASQAVKKG-PWKK 68
Db 29 TKIFINNEWHDSVSGKKFEVFNPAKEEKIC-EVAEGDKADIDKAVKAARKAFELGSPWRT 87
QY 69 FTGAQRAACMLKFADLAEKNAEKLARLESPTGRPVSMITHEDIPNMVSFRYYAGNADK 128

Db 88 MDASERGRLLNLADLVERDLTLATMEAIDGGKLFSTAYLMDLGACIKTIRYCAGWADK 147
QY 129 IAGKTFEPD-NGKPNWRYEPMGVCAGIASWNATFLYVGWKIAPALAAAGCSFIFKASEKSP 187
Db 148 IHGRTVPMDCGNFFFTTRHEPVGCGQIIPWNFPLVMEIWKIAPALCCGNTVVVKPAEQTP 207
QY 188 LGVILGLAPLFAEAGFPVGVVQFLTGARVT-GEALASHMDIAKISFTRSVGGGRAVQKATL 246
Db 208 LSALYMGSLIKEAGFPVGVVNIVPFGPTAGAAISHHMDIDKVSFTGSTEVGKLIKEAAG 267
QY 247 KSNMKRVTLLELGEK-PTIVFNEAPLERQSGESAKDFSKFGQIWVPPSCLLVQWGNLAEK- 304
Db 268 KTNLKRVTLELGGKSPNIIFADADL-----DEAAEFAHIGLFYHQGCCIAGSRIFVEEP 322
QY 305 -FHGVRHSGSGGCQRLWGNLQNPLEPKRTHGPFVDKSOYDRVLGNIDVG-KDTAQLLTGVGR 362
Db 323 IYDEFVRRSIERAKKTYTLGDPDLLPGVQQGPQIDKEQFKILDIESGKKEGAKLECGGGP 382
QY 363 KGDGFAIEPTIFVNPKPQSGKIWFEEIFGPVLSIKTFKTEEEAIEIANDTTYGLASVIYT 422
Db 383 WGNKGYFIQPTVFSNVTDDMRIAKBEIFGPVQQIMKFKTIDEVIKRRANNTYGLAAAVFT 442
QY 423 KSLNRGLRVSSALETGGVSINFPFIPETQTPFGGMKQSGSGRELGEGLKAYLEPKTINI 482
Db 443 KDIDKALTFASALQAGTVWVNCYSAPSAQCPFGFGKMSGNGRELGEYGLQYEVKTVTI 502
QY 483 HV 484
Db 503 KI 504

RESULT 12

S74224
aldehyde dehydrogenase (NAD+) (EC 1.2.1.3) 2 - mouse
N:Alternate names: retinaldehyde-specific dehydrogenase
C:Species: Mus musculus (house mouse)
C:Date: 04-Dec-1997 #sequence_revision 12-Dec-1997 #text_change 20-Jun-2000
C:Accession: S74224
R:Zhao, D.; McCaffery, P.; Ivins, K.J.; Neve, R.L.; Hogan, P.; Chin, W.W.; Draeger, U
Eur. J. Biochem. 240, 15-22, 1996
A:Title: Molecular identification of a major retinoic-acid-synthesizing enzyme, a ret
A:Reference number: S74224; MUID:96390857
A:Accession: S74224
A:Molecule type: mRNA
A:Residues: 1-499 <ZHA>
A:Cross-references: EMBL:X99273; NID:g1430868; PIDN:CAA67666.1; PID:g1430869
A:Experimental source: strain C3H/He; cell type embryonal carcinoma; cell line p19 te
C:Genetics:
A:Gene: RALDH-2
C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
C:Keywords: NAD; oxidoreductase
F:57-321/Domain: aldehyde dehydrogenase homology <ALDD>
F:193-273/Domain: NAD binding #status predicted <NAD>
F:267,301/Active site: Glu, Cys #status predicted
F:454/Binding site: NAD (Cys) #status predicted

Query Match 32.8%; Score 836.5; DB 2; Length 499;
Best Local Similarity 40.7%; Pred. No. 7.7e-58;
Matches 196; Conservative 83; Mismatches 191; Indels 11; Gaps 10;

QY 10 SELFFINNEFVSSKGSRLTLTNPWDESTVATDVHANAADVSAVAASQAVKKG-PWKK 68
Db 19 TKIFINNEWQNSGSRVFPVCNPAATGEQVC-EVQEADKVDIDKAVQAARLAFSLGSVWRR 77
QY 69 FTGAQRAACMLKFADLAEKNAEKLARLESPTGRPVSMITHEDIPNMVSFRYYAGWADK 128
Db 78 MDASERGRLLDLADLVERDRATLATNESLNGCKPFLOAFVIDLOGVIKTLRYIAGWADK 137
QY 129 IAGKTFPEDNGKPNW-RYEPMGVCAGIASWNATFLYVGWKIAPALAAAGCSFIFKASEKSP 187
Db 138 IHGMTIPVDGDIYFTTRHEPIGVCGQIIPWNFPLMTFTWKIAPALCCGNTVVVKPAEQTP 197

.....

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2002, 07:58:35 ; Search time 24.91 Seconds
(without alignments)
136.751 Million cell updates/sec

Title: US-09-882-694A-5
 Perfect score: 2549
 Sequence: 1 MVLSPDVKSELFINNEFVS.....EEGLKAYLEPKTINIHVITE 487

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 40831 seqs, 6994805 residues

Total number of hits satisfying chosen parameters: 40831

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 s

Database :	Pending_Patents_AA_New:*
1:	/cgn2_6/ptodata/2/paa/pct_NEW_COMB.pcp:*
2:	/cgn2_6/ptodata/2/paa/US06_NEW_COMB.pcp:*
3:	/cgn2_6/ptodata/2/paa/US07_NEW_COMB.pcp:*
4:	/cgn2_6/ptodata/2/paa/US08_NEW_COMB.pcp:*
5:	/cgn2_6/ptodata/2/paa/US09_NEW_COMB.pcp:*
6:	/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pcp:*
7:	/cgn2_6/ptodata/2/paa/US60_NEW_COMB.pcp:*

SUMMARIES

Result No.	Score	Query %		Length	DB	ID	Description
		Match					
1	561.5	22.0		528	5	US-09-573-655B-127	Sequence 127, App
2	394	15.5		607	5	US-09-573-655B-1086	Sequence 1086, Ap
3	261	10.2		108	6	US-10-106-698-6298	Sequence 6298, Ap
4	233.5	9.2		1185	5	US-09-895-913A-246	Sequence 246, App
5	96.5	3.8		1292	5	US-09-573-655B-1629	Sequence 1629, Ap
6	96.5	3.8		1292	5	US-09-573-655B-2282	Sequence 2282, Ap
7	94	3.7		573	5	US-09-847-637B-7	Sequence 7, Appli
8	93	3.6		518	5	US-09-540-209B-6751	Sequence 6751, Ap
9	91.5	3.6		612	5	US-09-573-655B-852	Sequence 852, App
10	91.5	3.6		612	5	US-09-573-655B-1114	Sequence 1114, Ap
11	90	3.5		735	5	US-09-978-403A-74	Sequence 74, Appl
12	90	3.5		735	5	US-09-978-544A-74	Sequence 74, Appl
13	90	3.5		735	5	US-09-978-681A-74	Sequence 74, Appl
14	90	3.5		735	5	US-09-978-757A-74	Sequence 74, Appl
15	90	3.5		735	5	US-09-978-564A-74	Sequence 74, Appl
16	90	3.5		735	5	US-09-999-831A-74	Sequence 74, Appl
17	90	3.5		735	5	US-09-999-829A-74	Sequence 74, Appl
18	90	3.5		735	6	US-10-013-921A-74	Sequence 74, Appl
19	90	3.5		735	6	US-10-013-929A-74	Sequence 74, Appl
20	90	3.5		735	6	US-10-013-918A-74	Sequence 74, Appl
21	90	3.5		735	6	US-10-017-082A-74	Sequence 74, Appl
22	88	3.5		573	5	US-09-847-637B-8	Sequence 8, Appli
23	86	3.4		529	5	US-09-540-209B-9534	Sequence 9534, Ap
24	85	3.3		395	5	US-09-540-209B-9876	Sequence 9876, Ap
25	85	3.3		575	6	US-10-108-605-55	Sequence 55, Appl
26	84.5	3.3		402	6	US-10-036-507-14	Sequence 14, Appl

ALIGNMENTS

```

RESULT      1
US-09-573-655B-127
; Sequence 127, Application US/09573655B
; GENERAL INFORMATION:
; APPLICANT: SOLOVYEV, Victor and TROUKHAN, M
; TITLE OF INVENTION: Sequence-Determined DN
; TITLE OF INVENTION: Thereby
; FILE REFERENCE: 2750-0876P
; CURRENT APPLICATION NUMBER: US/09/573,655B
; CURRENT FILING DATE: 2000-05-18
; NUMBER OF SEQ ID NOS: 3281
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 127
; LENGTH: 528
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-573-655B-127

```

Query Match 22.0%; Score 561.5; DB 5; Length 528;
Best Local Similarity 32.2%; Pred. NO. 6.1e-43;
Matches 160; Conservative 82; Mismatches 202; Indels 53; Gaps 15;

QY	9	KSELF	INNEFVSSKGSERLTLTNPWDESTVATDVHVANAADVD	SAAASVQAVKKGPPWK	68
DB	52	RTQGLIGKWLDSYDNKTIKVNPNPATGEIIA-DVACMGTKETND	AIASSYEAF	T--WSR	108
QY	69	FTGAQRAACMLKFADLAEKNAEKLARLES	LPTGRPVSMITHFDIPNMVSFRIYAGWADK	128	
DB	109	LTAGERSKVLRRWYDLLIAHKEELQGLITLEQKPKLKAIG-EVAYGASFI	EYIAEAKR	167	
QY	129	IAGKTFEPEDNGKPNWR-----YEPMGVCAGIASNNATFLYVGWKI	APALAAAGCSFIFK	181	
DB	168	VYGDIIIP-----PNLSDRRLVLKQPVGVVGAITPNWPF	PLAMITRKVGPALASGCTVVVK	222	
QY	182	ASEKSPGLVIGLAPLFAEAGFPBGVVQFLTG-ARVTGEALASHMDIAKIS	ETRSVGGGRA	240	
DB	223	PSELTPLTALAAAEALQAGVPPGALNVVMGNAPEIGDALLTSPQVRKIT	FTGSTAVGKK	282	
QY	241	VKQATLKSMMKRVLTLEL-GEKPTIVFNEAPLE-RQSGESAKDFSKFGQI	WVPSPCLLVQW	298	
DB	283	L-MAAAAPTVKVKSLELGGNAPSIVFDDADLDVAVKGTAAKFRNSGQT	CVCANRVLVQD	341	
QY	299	G---NLAERF-----HGVRHGSFGGCQRWLGNPLPKRTHGPFV	DKSQYDRVLG	345	
DB	342	GIYDKFAEAFSEAVQKLEVGDGFRDGT-----TOGPLINDAAVQK	VET	384	
QY	346	NI-DVGKDTAQLLTGVGRKGDKGFAI-EPTIFVNP	KPGSKIMFEEIFGPVLSIKTFKTEE	403	

Db 385 FVQDAVSKGAKIIG-GKRHSLGMTFYEPTVIRDVSDNMIMSKEEIFGPVAPLIRFKTEE 443
QY 404 EAIEIANDTTYGLASVIYTKSLNRLRVSSALETGGVSNFPFIPETQTPFGMKQSGSG 463
Db 444 DAIRIANDTIAGLAAYIFTNSVQSRWVRFEALEYGLVGVNEGLISTEVAPEFGVQSGLG 503
QY 464 RELGEEGLKAYLEPKTI 480
Db 504 REGSKYGMDEYLEIKYV 520

RESULT 2
US-09-573-655B-1086
; Sequence 1086, Application US/09573655B
; GENERAL INFORMATION:
; APPLICANT: SOLOVYEV, Victor and TROUKHAN, Maxim
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-0876P
; CURRENT APPLICATION NUMBER: US/09/573,655B
; CURRENT FILING DATE: 2000-05-18
; NUMBER OF SEQ ID NOS: 3281
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1086
; LENGTH: 607
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-573-655B-1086

Query Match 15.5%; Score 394; DB 5; Length 607;
Best Local Similarity 29.6%; Pred. No. 7.7e-28;
Matches 144; Conservative 64; Mismatches 241; Indels 38; Gaps 15;

QY 14 INNEFVSSKGSERLTLTNPWDESTVATDVHVANAADVSAVASQAVKKGPWKKFTGAQ 73
Db 118 IGGSFVESQSSSFIDVINPATQE-VVSKVPLTTNEEFKAAVSAKQAFPL--WRNTPITT 174
QY 74 RAACMLKFADLAENAEKLARLESPTGRPVSMITHEDIPNMVSVFRYYAGWADKIAGKT 133
Db 175 RQRVMLKFQELIRKNMDKLANITTEQGTKLK-DSHGDI FRGLEVVEHACGMATLQMGY 233
QY 134 FPE-DNGKPNWRY-EPMGVCAGIASWNATFLYVGWKIAPALAAAGCSFIFKASPKSLGLV 191
Db 234 LPNVSGVDTYSIREPLGVCAGICPNFPNAMIPLWMPFVAVTCGNTFILKPSKDPGASV 293
QY 192 GLAPLFAEAGFPFGVQVFLTGARVTGEALASHMDIAKISFTRSVGGGRAVKQATLKSNMK 251
Db 294 ILAELAMEAGLPDGLNIVHGTNDTVNAICDDEDIRAVSFVGSNTAGMHI-YARAAAKGK 352
QY 252 RVTLELGEK-PTIVFNEAPLERQ-SGESAKDFSKFGQIWVPPSCLLV-----QW-GNLAE 303
Db 353 RIQSNMGAKNHGLVLPDANIDATLALLAAGFGAAGQRCMALSTVVFVGDASKWEDKLVE 412
QY 304 KPHG--VRHGSFGCCQWLQGNPLEPKRTHGPFVDKSDVRLGNIDVGKDTAQLLTGVG 361
Db 413 RAKALKVTCGS-----EPDADLGPVISOAKERICRLIQSGVDDGAKLLLDG 459
QY 362 RK-----GDKGFAIEPTIFVNPKPCKIWFEEIFGPVLSIKTFKTEEEAIEIANDTTYGL 416
Db 460 KDIVVPGYERGNFPGTILSGVTPDMECYKEEIFGPVLCVMQANSFDEAISIINKNKYGN 519
QY 417 ASVIYTKSLNRLRVSSALETGGVSNFPF-IPETQTPFGMKQSGSG--RELGEGLKA 473
Db 520 GAAIFTSSGAARKFQMDIEAGQIGINVIPVPLPFPSFTGNKASFAGDLNPFYKAGVDF 579
QY 474 YLEPKTI 480
Db 580 FTQIKTV 586

RESULT 3
US-10-106-698-6298

; Sequence 6298, Application US/10106698
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypep
; FILE REFERENCE: PA005PI
; CURRENT APPLICATION NUMBER: US/10/106.698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 6298
; LENGTH: 108
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-106-698-6298

Query Match 10.2%; Score 261; DB 6; Length 108;
Best Local Similarity 52.4%; Pred. No. 5.7e-17;
Matches 54; Conservative 18; Mismatches 31; Indels 0; Gaps 0;

QY 382 SKIWFEEIFGPVLSIKTFKTEEEAIEIANDTTYGLASVIYTKSLNRLRVSSALETGVS 441
Db 1 AKIAKEEIFGPVMQILKFKTIEVVGRANNSTYGLAAAVFTKDLDKANYLSQALQAGTVW 60
QY 442 INEPPFIPETQTPFGMKQSGSGRELGEGLKAYLEPKTINIHV 484
Db 61 VNCYDVFGAQSPFGGYKMSGSGRELGEYGLQAYTEVKTVTVKV 103

RESULT 4
US-09-895-913A-246
; Sequence 246, Application US/09895913A
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean Francois
; APPLICANT: Oomen, Raymond P.
; TITLE OF INVENTION: Identification of polynucleotides
; TITLE OF INVENTION: Encoding Novel Helicobacter Polypeptides in the Helicobacter
; TITLE OF INVENTION: Genome
; FILE REFERENCE: 06132/043002
; CURRENT APPLICATION NUMBER: US/09/895.913A
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 08/881,227
; PRIOR FILING DATE: 1997-06-24
; NUMBER OF SEQ ID NOS: 368
; SOFTWARE: FastSeq for Windows version 4.0
; SEQ ID NO 246
; LENGTH: 1185
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-895-913A-246

Query Match 9.2%; Score 233.5; DB 5; Length 1185;
Best Local Similarity 26.1%; Pred. No. 4.8e-13;
Matches 117; Conservative 64; Mismatches 214; Indels 53; Gaps 18;

QY 42 VHANAADVDSAVAASVQAVKKGPWKKFTGAQRAACMLKFADLAENAEKLARLESPTG 101
Db 538 VHLADKEAILKALEVAKSDKSRFSQKSF--EIHALMSQTAQLFRERRGDLIGISALEVG 595
QY 102 RPYSMITHFDIPNMVSVFRYYAGWADKIAGKTFPEDNGKPNWRYEPMGVCAGIASWN-AT 160
Db 596 KTFAE-TDAEVSEIDPLEFY-----PYSLRVLQEONTKT--QFTPKGVGVVIAPWNFPV 647

QY	161	FLYVGWKIAPALAAAGCSFIFKASEKSPLGVLGLAPLFAEAGFPVGVQFLTGARVTGEAL	220
Dd	648	GISVG-TIAAPLATGNRIYKPPSSLSVGTGYKLCECFWDAGVPRDALIYLPS---KGSDI	703
QY	221	ASH-MDIAKTSFTRSVGGGRAVKQATLKSNMKRVITEL-----GEKPTIVNEAPLERQS	274
Dd	704	SEHLRDESIOFA-ILTGGEDTAYKMLKAN--PTIALSAETGGKNATIVSKMADRDRD---	756
QY	275	GESAKD-----FSKFQGIWPPSCLLVQWGNLAEK-----FHGVRHGSFGCCQRWLGN	323
Dd	757	-QAIKNVTHSAFNSGQKCSATSLLVLEKEVEYEDENFKKLIDATLSLVG-----D	807
QY	324	PLEPKRTHGPFVDKSOYDRVLGNDIVGKDTAQLLTGVGRKGDKGFAIEPTIFVNPKPGSK	383
Dd	808	PDFFNKKIGALADKPN-EKVKAIDELKSYENYEIPVSFVNDNPLYMKPISKYCTKKKGF	866
QY	384	IWFEEIFGPVLSIKTFKTEEEAIEIANDTTYGLASVITYTKSLNRGLRVSSALETTGGVSN	443
Dd	867	THQTelfTPILSVMEAKDLDEAIEIANSTGYGLTSALESIDEREWYYLRIEAGNIYN	926
QY	444	FPIPE--TQTPFGGMKQS--GSGRELG	467
Dd	927	KPTTGAIVRQOPFGGVKKSAVGFRKVG	954

```

RESULT      5
US-09-573-655B-1629
; Sequence 1629, Application US/09573655B
; GENERAL INFORMATION:
; APPLICANT: SOLOVYEV, Victor and TROUKHAN, Maxim
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Poly
; TITLE OF INVENTION: Thereby
; FILE REFERENCE: 2750-0876P
; CURRENT APPLICATION NUMBER: US/09/573,655B
; CURRENT FILING DATE: 2000-05-18
; NUMBER OF SEQ ID NOS: 3281
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1629
; LENGTH: 1292
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-573-655B-1629

Query Match      3.8%; Score 96.5; DB 5; Length 1292;
Best Local Similarity 22.0%; Pred. No. 1;
Matches 118; Conservative 68; Mismatches 201; Indels 149; Gaps 30;

```

```

DB   1028 LTACSHGLVDKGLAYYDMIARDFGLEPAIEHCICVIDLLGRSGRLAEAFISKMPMKP 1087
      :||| ||| : ||| : : : | | | : : ||
QY   381 GSKIWFE-----EIFGPVLSIKTFKEEEAIEIANDTYGLASVIYTKSLNRGLRVSSAL 435
      :| | | : :| | | : :| | | : :| | | : :| | | : :| | | : :| | | :
DB   1088 NDLVWRSLLASCKIHGNLDRGKAENLSKLEPEDDSVYVLSSNMFA-TTGRWEDVENVR 1146
      :| | | : :| | | : :| | | : :| | | : :| | | : :| | | : :| | | :
QY   436 ETGGVSINFPFIPETQTPFGGMKQSGSGRELGEF----GLKAYLEPKTINIHNIE 487
      :| | | : :| | | : :| | | : :| | | : :| | | : :| | | : :| | | :
DB   1147 QKMGI-----FKNIKK-----KQACSWVKLKDKVSSFGIGDRTHPQTMEIYAKLE 1190
      :| | | : :| | | : :| | | : :| | | : :| | | : :| | | : :| | | :

RESULT        6
US-09-573-655B-2282
; Sequence 2282, Application US/09S73655B
; GENERAL INFORMATION:
; APPLICANT: SOLOWEV, Victor and TROUKHAN, Maxim
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptid
; FILE REFERENCE: 2750-0876P
; CURRENT APPLICATION NUMBER: US/09/573,655B
; CURRENT FILING DATE: 2000-05-18
; NUMBER OF SEQ ID NOS: 3281
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2282
; LENGTH: 1292
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-573-655B-2282
```

```

Query Match          3.8%; Score 96.5; DB 5; Length 1292;
Best Local Similarity 22.0%; pred. No. 1;
Matches 118; Conservative 68; Mismatches 201; Indels 149; Gaps 30;

QY 10 SELFIN---NEFVSSKG-----SERLTLTNPWDESTVATDVHVANAA---DVSASVA 55
Db 746 SGLFYNQIIGNALVSMYKIGEMSESRRLVLLQMP-RRDVAVWNAALIGGYAEDEDPDKALA 804
QY 56 ASVQAVKKGPKWKFF--TGAQRAACMLKFADLAEKNAEKLARLESPLTGRPVS---MITHF 110
Db 805 AFQTRMRVEGVSSNYITVSVLSACULP-GDILLER-----GKPLHAYIVSAGF 850
QY 111 DIPNMV--SVFRYYAGWADKIACTFPEDNGKPNRWYEPMGVCAGIASWNAATFLYVGWKI 168
Db 851 ESDEHVKNLSITMYAKCGDLSSQDL--FNGLDN-----RNIIITWNA----- 890
QY 169 APALAAGCSFIFKASEKSPLGLVLGLAPLFAEAGFPVGVQFELTGARVTGEALASHMDIAK 228
Db 891 --MLAAN-----AHGHGEEVLKLVSKMRSFGVSLD----- 919
QY 229 ISFTRSVGGRAVKQATLK--SNMKRVTLLELG-EKPTIVFNEAPLERQSGESAKDFSKFG 285
Db 920 -QFSFSEGLSAAKLAVALLEEGQLHGLAVKGLGFEHDSFIFNAA-----ADMYSKCG 969
QY 286 QI-----WVPDSC--LLVQWGNLAEKPHGVHRHGSFGG-CQRW-----LGQNP-----L 325
Db 970 EIGEVVVKMLPSPVNRSLPSWNILISALG--RHGYFEEVCATFHEMLEMGIKPGHVTFVSL 1027
QY 326 EPKRTHGPFDK--SQYDRVLGNIDVGKDTAQLLTGVGRKDGKGAIEPTIFVNP---KP 380
Db 1028 LTACSHGGLVDKGLAYYDMIARDFGLPEAIEHCICVIDLLGRSGRLAEAEETFSKMPMKP 1087
QY 381 GSKIWFE-----EIFGPVLSIKTFKTEEEAIEANDTTYGLASVIYTKSLNRGLRVSSAL 435
Db 1088 NDLVWRSLASCKIHGNLDGRKAENLSKLEPEDDSVYVLSSNMFA-TTGRWEDVENVR 1146
QY 436 ETGGVSINFPIPETQTPFGGMKQSGSGRELGEE---GLKAYLEPKTINIHVNTE 487
Db 1147 QKMG---FKNIKK-----KQACSWVKLKDQVSSFGIGDRTHPQTMELIYAKLE 1190

RESULT 7

```


US-09-573-655B-1114

; Sequence 1114, Application US/09573655B
; GENERAL INFORMATION:
; APPLICANT: SOLOVYEV, Victor and TROUKHAN, Maxim
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: P2630PLC17
; CURRENT APPLICATION NUMBER: US/09/573,655B
; NUMBER OF SEQ ID NOS: 3281
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1114
; LENGTH: 612
; TYPE: PRP
; ORGANISM: Arabidopsis thaliana
US-09-573-655B-1114

Query Match 3.6%; Score 91.5; DB 5; Length 612;
Best Local Similarity 19.7%; Pred. No. 1;
Matches 88; Conservative 70; Mismatches 168; Indels 121; Gaps 20;

QY 4 SPDEKSEFINNEFVSCKSERLTLNPNWDESTATDVHANAADVDVAASQA-V-K 62
 | :
Db 180 SDEVNVKSVEGDEDGEDEDIVIVEKPVEERTDIDI-----ANVMEEAMQMPLGM 234
 | :
QY 63 KGPWKFTGAQRAA--CMLK----FADLAENAKNAELARLESPTGRPVSMITH---FDI 112
 | :
Db 235 YFPASEYTMMKLATRCYISEVLKTFALE-----HPLTNVENNYFMETH 278
 | :
QY 113 PNMSVFRIYYAGWADKIAG-----KTFPEONGKNPWRYEPMGVCAIASWNATFLYVGW 166
 | :
Db 279 PSFKHIHLPSGYTHKLMRMWMFLRTTSIEKKKEVW-YFKTGV----- 321
 | :
QY 167 KIAPALAAAGCSFIKFASKESPLGLVLGLAPLFAEAAGFFPGVVQLTGARY---TGEALAS 222
 | :
Db 322 -----IRREDVKTKL-MEMEPARSKDRLRMAVLYFLTSSIVVPKTGER-AS 366
 | :
QY 223 HMD-----IAKISFRSVGGGRAVKQATLSNMKRVTLELGEEKTIIVFNPAERQSCE 276
 | :
Db 367 PIDDFCVRAASDLTFCTFPNGKYSFYEMLKSIHTLDHFNGVVPNI---QSP----- 416
 | :
QY 277 SAKDESFGGIW-VPPSCILLVOW-----GNLAEKHGVHRHGSGGCORWLGNONPLEPKR 329
 | :
Db 417 -----WPVPGFCVPLEFLAFELAIPSLRRERFIEEGAHACPRMCKVN---FKR 462
 | :
QY 330 THGPFVDSQYDRVLGNIDV-----GKDTAQL--LTGVGRKGDKGFATEIPTIFVNP 379
 | :
Db 463 TEMKRFTLIQHNLHVLTTEVIESIIREKAEEVPLLAEITGVDEDDVDKHDVVSDSWMKRLG 522
 | :
QY 380 PGSKINFEETIGFVLSIKTFKTEEAI 406
 | :
Db 523 QRREIREEVEYNEDVHARMKAPNEEEV 549
 | :

RESULT 11
US-09-978-403A-74
; Sequence 74, Application US/09978403A
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Baker Kevin p.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan
; APPLICANT: Ferrara, Napoleon
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher

: APPLICANT: Gurney, Austin L.
: APPLICANT: Hillan, Kenneth J.
: APPLICANT: Kljavin, Ivar J.
: APPLICANT: Kuo, Sophia S.
: APPLICANT: Napier, Mary A.
: APPLICANT: Pan, James;
: APPLICANT: Paoni, Nicholas F.
: APPLICANT: Roy, Margaret Ann
: APPLICANT: Shelton, David L.
: APPLICANT: Stewart, Timothy A.
: APPLICANT: Tomas, Daniel
: APPLICANT: Williams, P. Mickey
: APPLICANT: Wood, William I.
: TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic Acids Encoding the Same
: FILE REFERENCE: P2630PLC17
: CURRENT APPLICATION NUMBER: US/09/978,403A
: CURRENT FILING DATE: 2002-03-19
: PRIOR APPLICATION NUMBER: 09/918585
: PRIOR FILING DATE: 2001-07-30
: PRIOR APPLICATION NUMBER: 60/062250
: PRIOR FILING DATE: 1997-10-17
: PRIOR APPLICATION NUMBER: 60/064249
: PRIOR FILING DATE: 1997-11-03
: PRIOR APPLICATION NUMBER: 60/065311
: PRIOR FILING DATE: 1997-11-13
: PRIOR APPLICATION NUMBER: 60/066364
: PRIOR FILING DATE: 1997-11-21
: PRIOR APPLICATION NUMBER: 60/077450
: PRIOR FILING DATE: 1998-03-10
: PRIOR APPLICATION NUMBER: 60/077632
: PRIOR FILING DATE: 1998-03-11
: PRIOR APPLICATION NUMBER: 60/077641
: PRIOR FILING DATE: 1998-03-11
: PRIOR APPLICATION NUMBER: 60/077649
: PRIOR FILING DATE: 1998-03-11
: PRIOR APPLICATION NUMBER: 60/077791
: PRIOR FILING DATE: 1998-03-12
: PRIOR APPLICATION NUMBER: 60/078004
: PRIOR FILING DATE: 1998-03-13
: PRIOR APPLICATION NUMBER: 60/078886
: PRIOR FILING DATE: 1998-03-20
: PRIOR APPLICATION NUMBER: 60/078936
: PRIOR FILING DATE: 1998-03-20
: PRIOR APPLICATION NUMBER: 60/078910
: PRIOR FILING DATE: 1998-03-20
: PRIOR APPLICATION NUMBER: 60/078939
: PRIOR FILING DATE: 1998-03-20
: PRIOR APPLICATION NUMBER: 60/079294
: PRIOR FILING DATE: 1998-03-25
: PRIOR APPLICATION NUMBER: 60/079656
: PRIOR FILING DATE: 1998-03-26
: PRIOR APPLICATION NUMBER: 60/079664
: PRIOR FILING DATE: 1998-03-27
: PRIOR APPLICATION NUMBER: 60/079689
: PRIOR FILING DATE: 1998-03-27
: PRIOR APPLICATION NUMBER: 60/079663
: PRIOR FILING DATE: 1998-03-27
: PRIOR APPLICATION NUMBER: 60/079728
: PRIOR FILING DATE: 1998-03-27
: PRIOR APPLICATION NUMBER: 60/079786
: PRIOR FILING DATE: 1998-03-27
: PRIOR APPLICATION NUMBER: 60/079920
: PRIOR FILING DATE: 1998-03-30
: PRIOR APPLICATION NUMBER: 60/079923
: PRIOR FILING DATE: 1998-03-30
: PRIOR APPLICATION NUMBER: 60/080105
: PRIOR FILING DATE: 1998-03-31
: PRIOR APPLICATION NUMBER: 60/080107
: PRIOR FILING DATE: 1998-03-31
: PRIOR APPLICATION NUMBER: 60/080165
: PRIOR FILING DATE: 1998-03-31
: PRIOR APPLICATION NUMBER: 60/080194

APPLICANT: Ferrara, Napoleon
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Kljavin, Ivar J.
APPLICANT: Kuo, Sophia S.
APPLICANT: Napier, Mary A.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Shelton, David L.
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: P2630P1C13
CURRENT APPLICATION NUMBER: US/09/978,544A
CURRENT FILING DATE: 2002-03-19
PRIOR APPLICATION NUMBER: 09/918585
PRIOR FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/064249
PRIOR FILING DATE: 1997-11-03
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066364
PRIOR FILING DATE: 1997-11-21
PRIOR APPLICATION NUMBER: 60/077450
PRIOR FILING DATE: 1998-03-10
PRIOR APPLICATION NUMBER: 60/077632
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077641
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077649
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077791
PRIOR FILING DATE: 1998-03-12
PRIOR APPLICATION NUMBER: 60/078004
PRIOR FILING DATE: 1998-03-13
PRIOR APPLICATION NUMBER: 60/078886
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078936
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078939
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/079294
PRIOR FILING DATE: 1998-03-25
PRIOR APPLICATION NUMBER: 60/079656
PRIOR FILING DATE: 1998-03-26
PRIOR APPLICATION NUMBER: 60/079664
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079689
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079663
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079728
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079786
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079920
PRIOR FILING DATE: 1998-03-30
PRIOR APPLICATION NUMBER: 60/079923
PRIOR FILING DATE: 1998-03-30
PRIOR APPLICATION NUMBER: 60/080105
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080107
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080165
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080194
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080327
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080328
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080333
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080334
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/081070
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081049
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081071
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081195
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081203
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081229
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081955
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081817
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081819
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081952
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081838
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/082568
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082569
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082704
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082804
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082700
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082797
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082796
PRIOR FILING DATE: 1998-04-23
PRIOR APPLICATION NUMBER: 60/083336
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/083392
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083495
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083496
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083499
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083545
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083554
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083558
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083559

; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079663
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079728
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079786
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079920
; PRIOR FILING DATE: 1998-03-30
; PRIOR APPLICATION NUMBER: 60/079923
; PRIOR FILING DATE: 1998-03-30
; PRIOR APPLICATION NUMBER: 60/080105
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080107
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080165
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080194
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080327
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080328
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080333
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080334
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/081070
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081049
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081071
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081195
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081203
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: 60/081229
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: 60/081955
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081817
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081819
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081952
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081838
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/082568
; PRIOR FILING DATE: 1998-04-21
; PRIOR APPLICATION NUMBER: 60/082569
; PRIOR FILING DATE: 1998-04-21
; PRIOR APPLICATION NUMBER: 60/082704
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082804
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082700
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082797
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082796
; PRIOR FILING DATE: 1998-04-23
; PRIOR APPLICATION NUMBER: 60/083336
; PRIOR FILING DATE: 1998-04-27
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/083392
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083495
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083496
; PRIOR FILING DATE: 1998-04-29

; PRIOR APPLICATION NUMBER: 60/083499
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083545
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083554
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083558
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083559
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083500
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083742
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: 60/084366
; PRIOR FILING DATE: 1998-05-05
; PRIOR APPLICATION NUMBER: 60/084414
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 60/084441
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 60/084637
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084639
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084640
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084598
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084627
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084643
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/085339
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085338
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085323
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085582
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085700
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085689
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085579
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085580
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085573
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085704
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085697
; PRIOR FILING DATE: 1998-05-15

Query Match 3.5%; Score 90; DB 5; Length 735;

Best Local Similarity 21.3%; Pred. No. 1.8;

Matches 63; Conservative 32; Mismatches 75; Indels 126; Gaps 14;

QY 239 RAVKQATLKSNMKRVTLGEGKPTIVFNEAPLEROSGESAKDFSFGQIWVPPSCLLVQW 298

Db 200 RHKRETLKAT-KYVEL-----VIVADNREFQRQ-----GKDLEKVKQ-----RLIEI 241

QY 299 GNLAKEFH--GVRHGSFGGQCRW-----LGQNP-----LEPKRTHGPF 334

Db 242 ANHVDKfYRPLNIRI-VLVGVEVWMDMKCSVSQDPFTSLJHEFLDWRKMKLLPRKSH--- 297

QY 335 VDKSQYDRVLGNIDVGKDTAQLLTGVGRKGDK-GFAIEPTIFVNPKPGSKIWFEEIFGPV 393

Db 298 -----DNAQLVSGVYFQCTIGMA-----PI 318

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2002, 07:55:18 ; Search time 389.72 Seconds
(without alignments)
439.840 Million cell updates/sec

Title: US-09-882-694A-5
Perfect score: 2549
Sequence: 1 MVLSPEYKSELFNNFVS.....EEGLKAYLEPKYTIHVNTIE 487

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 3502263 seqs, 351980561 residues
Total number of hits satisfying chosen parameters: 3502263

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_AA_Main:
1: /cgn2_6/ptodata/2/paa/PCTUS_COMB.pep:*
2: /cgn2_6/ptodata/2/paa/US06_COMB.pep:*
3: /cgn2_6/ptodata/2/paa/US07_COMB.pep:*
4: /cgn2_6/ptodata/2/paa/US080_COMB.pep:*
5: /cgn2_6/ptodata/2/paa/US081_COMB.pep:*
6: /cgn2_6/ptodata/2/paa/US082_COMB.pep:*
7: /cgn2_6/ptodata/2/paa/US083_COMB.pep:*
8: /cgn2_6/ptodata/2/paa/US084_COMB.pep:*
9: /cgn2_6/ptodata/2/paa/US085_COMB.pep:*
10: /cgn2_6/ptodata/2/paa/US086_COMB.pep:*
11: /cgn2_6/ptodata/2/paa/US087_COMB.pep:*
12: /cgn2_6/ptodata/2/paa/US088_COMB.pep:*
13: /cgn2_6/ptodata/2/paa/US089_COMB.pep:*
14: /cgn2_6/ptodata/2/paa/US090_COMB.pep:*
15: /cgn2_6/ptodata/2/paa/US091_COMB.pep:*
16: /cgn2_6/ptodata/2/paa/US092_COMB.pep:*
17: /cgn2_6/ptodata/2/paa/US093_COMB.pep:*
18: /cgn2_6/ptodata/2/paa/US094_COMB.pep:*
19: /cgn2_6/ptodata/2/paa/US095_COMB.pep:*
20: /cgn2_6/ptodata/2/paa/US096_COMB.pep:*
21: /cgn2_6/ptodata/2/paa/US097_COMB.pep:*
22: /cgn2_6/ptodata/2/paa/US098_COMB.pep:*
23: /cgn2_6/ptodata/2/paa/US099_COMB.pep:*
24: /cgn2_6/ptodata/2/paa/US100_COMB.pep:*
25: /cgn2_6/ptodata/2/paa/US101_COMB.pep:*
26: /cgn2_6/ptodata/2/paa/US60_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2549	100.0	487	17	US-09-351-224-5	Sequence 5, Appli
2	2549	100.0	487	17	US-09-351-224E-5	Sequence 5, Appli
3	2549	100.0	487	17	US-09-351-823-5	Sequence 5, Appli
4	2549	100.0	487	20	US-09-677-488-5	Sequence 5, Appli
5	2549	100.0	487	20	US-09-677-488A-5	Sequence 5, Appli
6	2549	100.0	487	20	US-09-677-682-5	Sequence 5, Appli
7	2549	100.0	487	20	US-09-677-682A-5	Sequence 5, Appli

8	2549	100.0	487	20	US-09-677-682B-5	Sequence 5, Appli
9	2549	100.0	487	22	US-09-882-694-5	Sequence 5, Appli
10	2549	100.0	487	22	US-09-882-694A-5	Sequence 5, Appli
11	886.5	34.8	559	21	US-09-760-443-1406	Sequence 1406, Ap
12	886.5	34.8	559	21	US-09-760-475-3429	Sequence 3429, Ap
13	882.5	34.6	517	1	PCT-US01-17253-3	Sequence 3, Appli
14	882.5	34.6	517	19	US-09-538-092-1047	Sequence 1047, Ap
15	868	34.1	497	10	US-08-612-853-10	Sequence 10, Appl
16	866	34.0	496	10	US-08-612-853-2	Sequence 2, Appli
17	866	34.0	496	10	US-08-612-855-3	Sequence 3, Appli
18	866	34.0	496	22	US-09-847-208-59	Sequence 59, Appl
19	866	34.0	496	22	US-09-847-208B-59	Sequence 59, Appl
20	864	33.9	497	10	US-08-612-853-11	Sequence 11, Appl
21	856.5	33.6	500	10	US-08-612-853-14	Sequence 14, Appl
22	852.5	33.4	500	10	US-08-612-853-15	Sequence 15, Appl
23	843.5	33.1	500	19	US-09-538-092-831	Sequence 831, App
24	840.5	33.0	517	19	US-09-538-092-851	Sequence 851, App
25	838.5	32.9	500	10	US-08-612-853-16	Sequence 16, Appl
26	836.5	32.8	499	6	US-08-238-818-2	Sequence 2, Appli
27	825.5	32.4	520	10	US-08-612-853-12	Sequence 12, Appl
28	824.5	32.3	512	19	US-09-538-092-1139	Sequence 1139, Ap
29	824.5	32.3	512	23	US-09-961-403-12	Sequence 12, Appl
30	824.5	32.3	515	7	US-08-366-210B-2	Sequence 2, Appli
31	824.5	32.3	529	1	PCT-US01-08631-36936	Sequence 36936, A
32	820.5	32.2	515	7	US-08-366-210B-4	Sequence 4, Appli
33	818.5	32.1	521	7	US-08-347-326A-2	Sequence 2, Appli
34	818.5	32.1	521	7	US-08-347-326B-2	Sequence 2, Appli
35	815.5	32.0	495	10	US-08-612-855-2	Sequence 2, Appli
36	815.5	32.0	495	22	US-09-847-208-11	Sequence 11, Appl
37	815.5	32.0	495	22	US-09-847-208B-11	Sequence 11, Appl
38	815.5	32.0	496	10	US-08-612-853-9	Sequence 9, Appli
39	808.5	31.7	500	10	US-08-612-853-13	Sequence 13, Appl
40	787	30.9	519	21	US-09-760-443-1247	Sequence 1247, Ap
41	787	30.9	519	21	US-09-760-475-3423	Sequence 3423, Ap
42	781.5	30.7	501	17	US-09-344-882-22	Sequence 22, Appl
43	781	30.6	538	17	US-09-344-882-20	Sequence 20, Appl
44	781	30.6	538	21	US-09-708-427-12049	Sequence 12049, A
45	779	30.6	511	21	US-09-708-427-55325	Sequence 55325, A

ALIGNMENTS

RESULT 1
US-09-351-224-5
; Sequence 5, Application US/09351224
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jon
; TITLE OF INVENTION: Compositions and Methods for Fummonisin Detoxification
; FILE REFERENCE: 5718-111
; CURRENT APPLICATION NUMBER: US/09/351,224
; CURRENT FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 487
; TYPE: PRT
; ORGANISM: Exophiala spinifera
; FEATURE:
; OTHER INFORMATION: aldehyde dehydrogenase
US-09-351-224-5

Query Match	100.0%;	Score 2549;	DB 17;	Length 487;						
Best Local Similarity	100.0%;	Pred. No. 3.4e-257;								
Matches 487;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;						
QY	1	MVLSPEYKSELFNNFVS	KGSR	LTLTNPWDESVATDVHVNADVD	SAVASVQA 60					
Db	1	MVLSPEYKSELFNNFVS	KGSR	LTLTNPWDESVATDVHVNADVD	SAVASVQA 60					
QY	61	VKKGPWKFTGAORACMLK	FEAD	LAEKNAEK	LRLES	SLPTGR	VPVSMITH	FDIP	NMVS	VFR 120

Db 61 VKKGPWKFTGAQRAACMLKFADLAEKNAEKLARLESPTGRPVSMITHFDIPNMVSVER 120
QY 121 YYAGWADKIAGKTFPEDNGKPNWRYEPMGVCAGIASWNATFLYVGWKIAPALAAAGCSFIF 180
Db 121 YYAGWADKIAGKTFPEDNGKPNWRYEPMGVCAGIASWNATFLYVGWKIAPALAAAGCSFIF 180
QY 181 KASEKSPGLVGLAPLFAEAGFPPGVVQFLTGARVTGEALASHMDIAKISFTRSVGGGRA 240
Db 181 KASEKSPGLVGLAPLFAEAGFPPGVVQFLTGARVTGEALASHMDIAKISFTRSVGGGRA 240
QY 241 VKQATLKSNNMKRVTLLEGEKPTIVFNEAPLERQSGESAKDFSKFGQIWWPPSCLLVQWGN 300
Db 241 VKQATLKSNNMKRVTLLEGEKPTIVFNEAPLERQSGESAKDFSKFGQIWWPPSCLLVQWGN 300
QY 301 LAEKFHGVRHGSFGGQCQWLGNQNPLEPKRTHGPFVDKSOYDRVLGNIDVGKDTAQLLTGV 360
Db 301 LAEKFHGVRHGSFGGQCQWLGNQNPLEPKRTHGPFVDKSOYDRVLGNIDVGKDTAQLLTGV 360
QY 361 GRKGDKGFAIEPTIFVNPCKPGSKIWFEEIFGVPVLSIKTFKTEEEAIEIANDTTYGLASVI 420
Db 361 GRKGDKGFAIEPTIFVNPCKPGSKIWFEEIFGVPVLSIKTFKTEEEAIEIANDTTYGLASVI 420
QY 421 YTKSLNRLGRVSSALETGGVSINFPFIPETQTPFGMKQSGSGRELGEGLKAYLEPKTI 480
Db 421 YTKSLNRLGRVSSALETGGVSINFPFIPETQTPFGMKQSGSGRELGEGLKAYLEPKTI 480
QY 481 NIHVNIE 487
Db 481 NIHVNIE 487

RESULT 2
US-09-351-224E-5
; Sequence 5, Application US/09351224E
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; TITLE OF INVENTION: Detoxification
; FILE REFERENCE: 5718-111
; CURRENT APPLICATION NUMBER: US/09/351,224E
; CURRENT FILING DATE: 1999-07-12
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 487
; TYPE: PRT
; ORGANISM: Exophiala spinifera
us-09-351-224E-5

Query Match 100.0%; Score 2549; DB 17; Length 487;
Best Local Similarity 100.0%; Pred. No. 3.4e-257;
Matches 487; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVLSPDEYKSELFNNFVSSKGSERLTLTNPWDESTVATDVHANAADVDSAVAASVQA 60
Db 1 MVLSPDEYKSELFNNFVSSKGSERLTLTNPWDESTVATDVHANAADVDSAVAASVQA 60
QY 61 VKKGPWKFTGAQRAACMLKFADLAEKNAEKLARLESPTGRPVSMITHFDIPNMVSVER 120
Db 61 VKKGPWKFTGAQRAACMLKFADLAEKNAEKLARLESPTGRPVSMITHFDIPNMVSVER 120
QY 121 YYAGWADKIAGKTFPEDNGKPNWRYEPMGVCAGIASWNATFLYVGWKIAPALAAAGCSFIF 180
Db 121 YYAGWADKIAGKTFPEDNGKPNWRYEPMGVCAGIASWNATFLYVGWKIAPALAAAGCSFIF 180
QY 181 KASEKSPGLVGLAPLFAEAGFPPGVVQFLTGARVTGEALASHMDIAKISFTRSVGGGRA 240
Db 181 KASEKSPGLVGLAPLFAEAGFPPGVVQFLTGARVTGEALASHMDIAKISFTRSVGGGRA 240

QY 241 VKQATLKSNNMKRVTLLEGEKPTIVFNEAPLERQSGESAKDFSKFGQIWWPPSCLLVQWGN 300
Db 241 VKQATLKSNNMKRVTLLEGEKPTIVFNEAPLERQSGESAKDFSKFGQIWWPPSCLLVQWGN 300
QY 301 LAEKFHGVRHGSFGGQCQWLGNQNPLEPKRTHGPFVDKSOYDRVLGNIDVGKDTAQLLTGV 360
Db 301 LAEKFHGVRHGSFGGQCQWLGNQNPLEPKRTHGPFVDKSOYDRVLGNIDVGKDTAQLLTGV 360
QY 361 GRKGDKGFAIEPTIFVNPCKPGSKIWFEEIFGVPVLSIKTFKTEEEAIEIANDTTYGLASVI 420
Db 361 GRKGDKGFAIEPTIFVNPCKPGSKIWFEEIFGVPVLSIKTFKTEEEAIEIANDTTYGLASVI 420
QY 421 YTKSLNRLGRVSSALETGGVSINFPFIPETQTPFGMKQSGSGRELGEGLKAYLEPKTI 480
Db 421 YTKSLNRLGRVSSALETGGVSINFPFIPETQTPFGMKQSGSGRELGEGLKAYLEPKTI 480
QY 481 NIHVNIE 487
Db 481 NIHVNIE 487

RESULT 3
US-09-351-823-5
; Sequence 5, Application US/09351823
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jon
; TITLE OF INVENTION: Compositions and Methods for Fumonisin Detoxification
; FILE REFERENCE: 5718-56
; CURRENT APPLICATION NUMBER: US/09/351,823
; CURRENT FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 487
; TYPE: PRT
; ORGANISM: Exophiala spinifera
; FEATURE:
; OTHER INFORMATION: aldehyde dehydrogenase
US-09-351-823-5

Query Match 100.0%; Score 2549; DB 17; Length 487;
Best Local Similarity 100.0%; Pred. No. 3.4e-257;
Matches 487; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVLSPDEYKSELFNNFVSSKGSERLTLTNPWDESTVATDVHANAADVDSAVAASVQA 60
Db 1 MVLSPDEYKSELFNNFVSSKGSERLTLTNPWDESTVATDVHANAADVDSAVAASVQA 60
QY 61 VKKGPWKFTGAQRAACMLKFADLAEKNAEKLARLESPTGRPVSMITHFDIPNMVSVER 120
Db 61 VKKGPWKFTGAQRAACMLKFADLAEKNAEKLARLESPTGRPVSMITHFDIPNMVSVER 120
QY 121 YYAGWADKIAGKTFPEDNGKPNWRYEPMGVCAGIASWNATFLYVGWKIAPALAAAGCSFIF 180
Db 121 YYAGWADKIAGKTFPEDNGKPNWRYEPMGVCAGIASWNATFLYVGWKIAPALAAAGCSFIF 180
QY 181 KASEKSPGLVGLAPLFAEAGFPPGVVQFLTGARVTGEALASHMDIAKISFTRSVGGGRA 240
Db 181 KASEKSPGLVGLAPLFAEAGFPPGVVQFLTGARVTGEALASHMDIAKISFTRSVGGGRA 240
QY 241 VKQATLKSNNMKRVTLLEGEKPTIVFNEAPLERQSGESAKDFSKFGQIWWPPSCLLVQWGN 300
Db 241 VKQATLKSNNMKRVTLLEGEKPTIVFNEAPLERQSGESAKDFSKFGQIWWPPSCLLVQWGN 300
QY 301 LAEKFHGVRHGSFGGQCQWLGNQNPLEPKRTHGPFVDKSOYDRVLGNIDVGKDTAQLLTGV 360
Db 301 LAEKFHGVRHGSFGGQCQWLGNQNPLEPKRTHGPFVDKSOYDRVLGNIDVGKDTAQLLTGV 360
QY 361 GRKGDKGFAIEPTIFVNPCKPGSKIWFEEIFGVPVLSIKTFKTEEEAIEIANDTTYGLASVI 420
Db 361 GRKGDKGFAIEPTIFVNPCKPGSKIWFEEIFGVPVLSIKTFKTEEEAIEIANDTTYGLASVI 420

FILE REFERENCE: 35718/204101 (5718-111B)
CURRENT APPLICATION NUMBER: US/09/677,682
CURRENT FILING DATE: 2000-10-02
PRIOR APPLICATION NUMBER: 09/351,224
PRIOR FILING DATE: 1999-07-12
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 5
LENGTH: 487
TYPE: PRT
ORGANISM: Exophiala spinifera
FEATURE:
OTHER INFORMATION: aldehyde dehydrogenase, translation of fully
OTHER INFORMATION: spliced cDNA
US-09-677-682-5

Query Match 100.0%; Score 2549; DB 20; Length 487;
Best Local Similarity 100.0%; Pred. No. 3.4e-257;
Matches 487; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MVLSPDEYKSELFNNFVSSKSGSERLTLTNPWDESTVATDVHANAADVDSAVAASVQA 60
Db 1 MVLSPDEYKSELFNNFVSSKSGSERLTLTNPWDESTVATDVHANAADVDSAVAASVQA 60
Qy 61 VKKGPWKFTGAQRAACMLKFADLAEKNAEKLAERLESPTGRPVSMITHFDIPNMVSVER 120
Db 61 VKKGPWKFTGAQRAACMLKFADLAEKNAEKLAERLESPTGRPVSMITHFDIPNMVSVER 120
Qy 121 YYAGWADKIAGKTFPEDNGKPNWRYEPMGVGCAGIASWNATFLYVGWKIAPALAAAGCSFIF 180
Db 121 YYAGWADKIAGKTFPEDNGKPNWRYEPMGVGCAGIASWNATFLYVGWKIAPALAAAGCSFIF 180
Qy 181 KASEKSPGLVGLAPLFAEAGFPVGVVQFLTGARVTGEALASHMDIAKISFTRSVGGGRA 240
Db 181 KASEKSPGLVGLAPLFAEAGFPVGVVQFLTGARVTGEALASHMDIAKISFTRSVGGGRA 240
Qy 241 VKQATLKSNNMKRVTLLELGEKPTIVFNEAPLERQSGESAKDFSKFGQIWVPPSCLLVQWGN 300
Db 241 VKQATLKSNNMKRVTLLELGEKPTIVFNEAPLERQSGESAKDFSKFGQIWVPPSCLLVQWGN 300
Qy 301 LAEKFHGVRHGSFGCCQWLGNQNPLEPKRTHGPFVDKSDYDRVLGNIDVGDVTAQLLTGV 360
Db 301 LAEKFHGVRHGSFGCCQWLGNQNPLEPKRTHGPFVDKSDYDRVLGNIDVGDVTAQLLTGV 360
Qy 361 GRKGDGFAIEPTIFVNPKPCKGKIWFEEIFGVPVLSIKTFKTEEEAIEIANDTTYGASVI 420
Db 361 GRKGDGFAIEPTIFVNPKPCKGKIWFEEIFGVPVLSIKTFKTEEEAIEIANDTTYGASVI 420
Qy 421 YTKSLNRLGRVSSALETGGVSINFPFIPETQTPFGGMKQSGSGRELGEGLKAYLEPKTI 480
Db 421 YTKSLNRLGRVSSALETGGVSINFPFIPETQTPFGGMKQSGSGRELGEGLKAYLEPKTI 480
Qy 481 NIHVNIE 487
Db 481 NIHVNIE 487

RESULT 7
US-09-677-682A-5
Sequence 5, Application US/09677682A
GENERAL INFORMATION:
APPLICANT: Duvick, Jon
APPLICANT: Maddox, Joyce
APPLICANT: Gilliam, Jacob
APPLICANT: Folkerts, Otto
APPLICANT: Crasta, Oswald R.
TITLE OF INVENTION: Compositions and Methods for Fumonisin
FILE REFERENCE: 35718/204101
CURRENT APPLICATION NUMBER: US/09/677,682A
CURRENT FILING DATE: 2000-10-02
PRIOR APPLICATION NUMBER: 09/351,224

PRIOR FILING DATE: 1999-07-12
NUMBER OF SEQ ID NOS: 11
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5
LENGTH: 487
TYPE: PRT
ORGANISM: Exophiala spinifera
US-09-677-682A-5

Query Match 100.0%; Score 2549; DB 20; Length 487;
Best Local Similarity 100.0%; Pred. No. 3.4e-257;
Matches 487; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MVLSPDEYKSELFNNFVSSKSGSERLTLTNPWDESTVATDVHANAADVDSAVAASVQA 60
Db 1 MVLSPDEYKSELFNNFVSSKSGSERLTLTNPWDESTVATDVHANAADVDSAVAASVQA 60
Qy 61 VKKGPWKFTGAQRAACMLKFADLAEKNAEKLAERLESPTGRPVSMITHFDIPNMVSVER 120
Db 61 VKKGPWKFTGAQRAACMLKFADLAEKNAEKLAERLESPTGRPVSMITHFDIPNMVSVER 120
Qy 121 YYAGWADKIAGKTFPEDNGKPNWRYEPMGVGCAGIASWNATFLYVGWKIAPALAAAGCSFIF 180
Db 121 YYAGWADKIAGKTFPEDNGKPNWRYEPMGVGCAGIASWNATFLYVGWKIAPALAAAGCSFIF 180
Qy 181 KASEKSPGLVGLAPLFAEAGFPVGVVQFLTGARVTGEALASHMDIAKISFTRSVGGGRA 240
Db 181 KASEKSPGLVGLAPLFAEAGFPVGVVQFLTGARVTGEALASHMDIAKISFTRSVGGGRA 240
Qy 241 VKQATLKSNNMKRVTLLELGEKPTIVFNEAPLERQSGESAKDFSKFGQIWVPPSCLLVQWGN 300
Db 241 VKQATLKSNNMKRVTLLELGEKPTIVFNEAPLERQSGESAKDFSKFGQIWVPPSCLLVQWGN 300
Qy 301 LAEKFHGVRHGSFGCCQWLGNQNPLEPKRTHGPFVDKSDYDRVLGNIDVGDVTAQLLTGV 360
Db 301 LAEKFHGVRHGSFGCCQWLGNQNPLEPKRTHGPFVDKSDYDRVLGNIDVGDVTAQLLTGV 360
Qy 361 GRKGDGFAIEPTIFVNPKPCKGKIWFEEIFGVPVLSIKTFKTEEEAIEIANDTTYGASVI 420
Db 361 GRKGDGFAIEPTIFVNPKPCKGKIWFEEIFGVPVLSIKTFKTEEEAIEIANDTTYGASVI 420
Qy 421 YTKSLNRLGRVSSALETGGVSINFPFIPETQTPFGGMKQSGSGRELGEGLKAYLEPKTI 480
Db 421 YTKSLNRLGRVSSALETGGVSINFPFIPETQTPFGGMKQSGSGRELGEGLKAYLEPKTI 480
Qy 481 NIHVNIE 487
Db 481 NIHVNIE 487

RESULT 8
US-09-677-682B-5
Sequence 5, Application US/09677682B
GENERAL INFORMATION:
APPLICANT: Duvick, Jon
APPLICANT: Maddox, Joyce
APPLICANT: Gilliam, Jacob
APPLICANT: Folkerts, Otto
APPLICANT: Crasta, Oswald R.
TITLE OF INVENTION: Compositions and Methods for Fumonisin
FILE REFERENCE: 35718/204101
CURRENT APPLICATION NUMBER: US/09/677,682B
CURRENT FILING DATE: 2000-10-02
PRIOR APPLICATION NUMBER: 09/351,224
PRIOR FILING DATE: 1999-07-12
NUMBER OF SEQ ID NOS: 11
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5
LENGTH: 487
TYPE: PRT
ORGANISM: Exophiala spinifera

US-09-677-682B-5

Query Match 100.0%; Score 2549; DB 20; Length 487;
Best Local Similarity 100.0%; Pred. No. 3.4e-257;
Matches 487; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVLSPDEYKSELFNNFVSSKSGSERLTLTNPWDESTVATDVHANAADVDSAVAASVQA 60
Db 1 MVLSPDEYKSELFNNFVSSKSGSERLTLTNPWDESTVATDVHANAADVDSAVAASVQA 60
QY 61 VKKGPWKFTGAQRAACMLKFADLAEKNAEKLARLESPTGRPVSMITHFDIPNMVSVER 120
Db 1 MVLSPDEYKSELFNNFVSSKSGSERLTLTNPWDESTVATDVHANAADVDSAVAASVQA 60
QY 61 VKKGPWKFTGAQRAACMLKFADLAEKNAEKLARLESPTGRPVSMITHFDIPNMVSVER 120
Db 61 VKKGPWKFTGAQRAACMLKFADLAEKNAEKLARLESPTGRPVSMITHFDIPNMVSVER 120
QY 121 YYAGWADKIAGKTFPEDNGKPNWRYEPMGVGCAGIASWNATFLYVGWKIAPALAAAGCSFIF 180
Db 121 YYAGWADKIAGKTFPEDNGKPNWRYEPMGVGCAGIASWNATFLYVGWKIAPALAAAGCSFIF 180
QY 181 KASEKSPLGVLGLAPLFAEAGFPVGVQFLTGARVTGEALASHMDIAKISFTRSVGGGRA 240
Db 181 KASEKSPLGVLGLAPLFAEAGFPVGVQFLTGARVTGEALASHMDIAKISFTRSVGGGRA 240
QY 241 VKQATLKSNNKRVTLLELGEKPTIVFNEAPLERQSGESAKDFSKFGQIWVPPSCLLVQWGN 300
Db 241 VKQATLKSNNKRVTLLELGEKPTIVFNEAPLERQSGESAKDFSKFGQIWVPPSCLLVQWGN 300
QY 301 LAEKFHGVRHSGFGGCQRLWQNLPLEKRTHPFVDKSYDRVLGNIDVGKDTAQLLTGV 360
Db 301 LAEKFHGVRHSGFGGCQRLWQNLPLEKRTHPFVDKSYDRVLGNIDVGKDTAQLLTGV 360
QY 361 GRKGDKGFAIEPTIFVNPKPCKGIWFEIEFGPVLSIKTFKTEEEAIEIANDTTYGLASVI 420
Db 361 GRKGDKGFAIEPTIFVNPKPCKGIWFEIEFGPVLSIKTFKTEEEAIEIANDTTYGLASVI 420
QY 421 YTKSLNRGLRVSSALETGGVSINFPFIPETQTPFGGMKQSGSGRELGEGLKAYLEPKTI 480
Db 421 YTKSLNRGLRVSSALETGGVSINFPFIPETQTPFGGMKQSGSGRELGEGLKAYLEPKTI 480
QY 481 NIHVNIE 487
Db 481 NIHVNIE 487

RESULT 9
US-09-882-694-5
; Sequence 5, Application US/09882694
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; FILE REFERENCE: 35718/208255
; CURRENT APPLICATION NUMBER: US/09/882,694
; CURRENT FILING DATE: 2001-06-12
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 487
; TYPE: PRT
; ORGANISM: Exophiala spinifera
US-09-882-694-5

Query Match 100.0%; Score 2549; DB 22; Length 487;
Best Local Similarity 100.0%; Pred. No. 3.4e-257;
Matches 487; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVLSPDEYKSELFNNFVSSKSGSERLTLTNPWDESTVATDVHANAADVDSAVAASVQA 60
Db 1 MVLSPDEYKSELFNNFVSSKSGSERLTLTNPWDESTVATDVHANAADVDSAVAASVQA 60
QY 61 VKKGPWKFTGAQRAACMLKFADLAEKNAEKLARLESPTGRPVSMITHFDIPNMVSVER 120
Db 61 VKKGPWKFTGAQRAACMLKFADLAEKNAEKLARLESPTGRPVSMITHFDIPNMVSVER 120
QY 121 YYAGWADKIAGKTFPEDNGKPNWRYEPMGVGCAGIASWNATFLYVGWKIAPALAAAGCSFIF 180
Db 121 YYAGWADKIAGKTFPEDNGKPNWRYEPMGVGCAGIASWNATFLYVGWKIAPALAAAGCSFIF 180
QY 181 KASEKSPLGVLGLAPLFAEAGFPVGVQFLTGARVTGEALASHMDIAKISFTRSVGGGRA 240
Db 181 KASEKSPLGVLGLAPLFAEAGFPVGVQFLTGARVTGEALASHMDIAKISFTRSVGGGRA 240
QY 241 VKQATLKSNNKRVTLLELGEKPTIVFNEAPLERQSGESAKDFSKFGQIWVPPSCLLVQWGN 300
Db 241 VKQATLKSNNKRVTLLELGEKPTIVFNEAPLERQSGESAKDFSKFGQIWVPPSCLLVQWGN 300
QY 301 LAEKFHGVRHSGFGGCQRLWQNLPLEKRTHPFVDKSYDRVLGNIDVGKDTAQLLTGV 360
Db 301 LAEKFHGVRHSGFGGCQRLWQNLPLEKRTHPFVDKSYDRVLGNIDVGKDTAQLLTGV 360
QY 361 GRKGDKGFAIEPTIFVNPKPCKGIWFEIEFGPVLSIKTFKTEEEAIEIANDTTYGLASVI 420
Db 361 GRKGDKGFAIEPTIFVNPKPCKGIWFEIEFGPVLSIKTFKTEEEAIEIANDTTYGLASVI 420
QY 421 YTKSLNRGLRVSSALETGGVSINFPFIPETQTPFGGMKQSGSGRELGEGLKAYLEPKTI 480
Db 421 YTKSLNRGLRVSSALETGGVSINFPFIPETQTPFGGMKQSGSGRELGEGLKAYLEPKTI 480
QY 481 NIHVNIE 487
Db 481 NIHVNIE 487

RESULT 10
US-09-882-694A-5
; Sequence 5, Application US/09882694A
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; FILE REFERENCE: 5718-111
; CURRENT APPLICATION NUMBER: US/09/882,694A
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 487
; TYPE: PRT
; ORGANISM: Exophiala spinifera
US-09-882-694A-5

Query Match 100.0%; Score 2549; DB 22; Length 487;
Best Local Similarity 100.0%; Pred. No. 3.4e-257;
Matches 487; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVLSPDEYKSELFNNFVSSKSGSERLTLTNPWDESTVATDVHANAADVDSAVAASVQA 60
Db 1 MVLSPDEYKSELFNNFVSSKSGSERLTLTNPWDESTVATDVHANAADVDSAVAASVQA 60
QY 61 VKKGPWKFTGAQRAACMLKFADLAEKNAEKLARLESPTGRPVSMITHFDIPNMVSVER 120
Db 61 VKKGPWKFTGAQRAACMLKFADLAEKNAEKLARLESPTGRPVSMITHFDIPNMVSVER 120


```
Db      248 VVMKVAEQTPLSALYLASLIKEAGPPGVNIIITGYGPTAGAAIAQHMDVDKVAFTGSTE 307
QY      237 GGRAVKQATLKSNMKRVTLLELGEK-PTIVNEAPLE---RQSGESAKDFSKEFGQIWWPPS 292
Db      308 VGHLIQKAAGDSNLKRVLTLELGGKSPSIVLADADMEHAVEQCHEAL--FFNMGQCCAGS 365
QY      293 CLLVQWG-----NLAKEFHGVRHGSFGGCQRWLGNPLEPKRTHGPFVVDKSOYDRVLG 345
Db      366 RTFVEESIYNEFLERTVEK-----AKORKVG-NPFELDTQQGPQVDKEQFERVLG 414
QY      346 NIDVG-KDTAQLLTGVGRKGDKGFAIEPTIFVNPKPCKGSKIWFEEIFGPVLSIKTFKTEEE 404
Db      415 YIQLGOKEGAKLLCGGERFGERGFFIKPTVFGGVQDDMRIAKEEIEFGPVQPLFKFKKIEE 474
QY      405 AIEIANDTTYGLASVIYTKSLNRLRVSSALETGGVSINFPFIPETQTPFGGMKQSGSGR 464
Db      475 VVERANNTRYGLAAAVFTRDLDKAMYFTQALQAGTVVWNTYNIVTCHTTPFGGFKESGNGR 534
QY      465 ELGEEGLKAYLEPKTINIHV 484
Db      535 ELGEDGLKAYTEVKTIVKV 554

RESULT 13
PCT-US01-17253-3
; Sequence 3, Application PC/TUS0117253
; GENERAL INFORMATION:
; APPLICANT: Genaissance Pharmaceuticals
; APPLICANT: Duda, Amy
; APPLICANT: Finkel, Kevin
; APPLICANT: Kazemi, Amir
; APPLICANT: Messer, Chad
; APPLICANT: Sanchis, Angela
; TITLE OF INVENTION: Haplotypes of the ALDH5 Gene
; FILE REFERENCE: MWH-0719PCT ALDH5
; CURRENT APPLICATION NUMBER: PCT/US01/17253
; CURRENT FILING DATE: 2001-05-29
; PRIOR APPLICATION NUMBER: 60/207,508
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 89
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 517
; TYPE: PRT
; ORGANISM: Homo sapien
PCT-US01-17253-3

Query Match      34.6%; Score 882.5; DB 1; Length 517;
Best Local Similarity 41.9%; Pred. No. 1.7e-82;
Matches 210; Conservative 80; Mismatches 176; Indels 35; Gaps 13;

QY      2 VLSPEYKSELFINNEFVSSKGSERLTLTNPWDESTVATDVHVA--NAADVDSAVAASVQ 59
Db      29 ILNPDIPYNQLFINNEWQDAVSKKTFTPTVNPTTGEVIG---HVAEGDRADVRAVKAARE 85
QY      60 AVKKG-PWKKFTGAQRAACMLKFADLAEKNAEKLARLESLEPTGRPVSMITHFDIPNMVSV 118
Db      86 AFRLCSPWRRMDASERGRLLNLLADLVERDRVYLASLETLDNGKPFQESYALDLDEVIKV 145
QY      119 FRYAGWADKIAGKTFPEDNGKP--NWRYEPMGVGCAGIASWNATFLYVGWKIAPALAAGC 176
Db      146 YRYFAGWADKWHGKTIPM-HGQHFCFTRHEPVGVCQGIIPWNFPLVMQGWKLAPALATGN 204
QY      177 SFIFKASEKSPLGLGLAPLFAEAGPPGVVQFLTGARVT-GEALASHMDIAKISFTRSV 235
Db      205 TVVMKVAEQTPLSALYLASLIKEAGPPGVVNIITGYGPTAGAAIAQHMDVDKVAFTGST 264
QY      236 GGRAVKQATLKSNMKRVTLLELGEK-PTIVNEAPLE---RQSGESAKDFSKEFGQIWWPP 291
Db      265 EVGHLIQKAAGDSNLKRVLTLELGGKSPSIVLADADMEHAVEQCHEAL--FFNMGQCCAG 322
QY      292 SCLLVQWG-----NLAKEFHGVRHGSFGGCQRWLGNPLEPKRTHGPFVVDKSOYDRVL 344
Db      323 RTFVEESIYNEFLERTVEK-----AKORKVG-NPFELDTQQGPQVDKEQFERVL 371
QY      345 GNIDVG-KDTAQLLTGVGRKGDKGFAIEPTIFVNPKPCKGSKIWFEEIFGPVLSIKTFKTEE 403
```

```
QY      292 SCLLVQWG-----NLAKEFHGVRHGSFGGCQRWLGNPLEPKRTHGPFVVDKSOYDRVL 344
Db      323 RTFVEESIYNEFLERTVEK-----AKORKVG-NPFELDTQQGPQVDKEQFERVL 371
QY      345 GNIDVG-KDTAQLLTGVGRKGDKGFAIEPTIFVNPKPCKGSKIWFEEIFGPVLSIKTFKTEE 403
Db      372 GYIQLGOKEGAKLLCGGERFGERGFFIKPTVFGGVQDDMRIAKEEIEFGPVQPLFKFKKIE 431
QY      404 EAIEIANDTTYGLASVIYTKSLNRLRVSSALETGGVSINFPFIPETQTPFGGMKQSGSG 463
Db      432 EVVERANNTRYGLAAAVFTRDLDKAMYFTQALQAGTVVWNTYNIVTCHTTPFGGFKESGNG 491
QY      464 RELGEEGLKAYLEPKTINIHV 484
Db      492 RELGEDGLKAYTEVKTIVKV 512

RESULT 14
US-09-538-092-1047
; Sequence 1047, Application US/09538092
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CurapatSeqformatter Version 0.9
; SEQ ID NO 1047
; LENGTH: 517
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number P30837
US-09-538-092-1047

Query Match      34.6%; Score 882.5; DB 19; Length 517;
Best Local Similarity 41.9%; Pred. No. 1.7e-82;
Matches 210; Conservative 80; Mismatches 176; Indels 35; Gaps 13;

QY      2 VLSPEYKSELFINNEFVSSKGSERLTLTNPWDESTVATDVHVA--NAADVDSAVAASVQ 59
Db      29 ILNPDIPYNQLFINNEWQDAVSKKTFTPTVNPTTGEVIG---HVAEGDRADVRAVKAARE 85
QY      60 AVKKG-PWKKFTGAQRAACMLKFADLAEKNAEKLARLESLEPTGRPVSMITHFDIPNMVSV 118
Db      86 AFRLCSPWRRMDASERGRLLNLLADLVERDRVYLASLETLDNGKPFQESYALDLDEVIKV 145
QY      119 FRYAGWADKIAGKTFPEDNGKP--NWRYEPMGVGCAGIASWNATFLYVGWKIAPALAAGC 176
Db      146 YRYFAGWADKWHGKTIPM-HGQHFCFTRHEPVGVCQGIIPWNFPLVMQGWKLAPALATGN 204
QY      177 SFIFKASEKSPLGLGLAPLFAEAGPPGVVQFLTGARVT-GEALASHMDIAKISFTRSV 235
Db      205 TVVMKVAEQTPLSALYLASLIKEAGPPGVVNIITGYGPTAGAAIAQHMDVDKVAFTGST 264
QY      236 GGRAVKQATLKSNMKRVTLLELGEK-PTIVNEAPLE---RQSGESAKDFSKEFGQIWWPP 291
Db      265 EVGHLIQKAAGDSNLKRVLTLELGGKSPSIVLADADMEHAVEQCHEAL--FFNMGQCCAG 322
QY      292 SCLLVQWG-----NLAKEFHGVRHGSFGGCQRWLGNPLEPKRTHGPFVVDKSOYDRVL 344
Db      323 RTFVEESIYNEFLERTVEK-----AKORKVG-NPFELDTQQGPQVDKEQFERVL 371
QY      345 GNIDVG-KDTAQLLTGVGRKGDKGFAIEPTIFVNPKPCKGSKIWFEEIFGPVLSIKTFKTEE 403
```

Db 372 GYIQLGQKEGAKLLCGGERGFFIKPTVFGVQDDMRIAKKEIFGPVQPLFKFKKIE 431
QY 404 EAIEIANDTYGLASVIYTKSLNRGLRVSSALETGGVSINFPFIPETQTPFGGMKQSGG 463
Db 432 EVVERANNTRYGLAAAVFTROLDKRAMYFTQALQAGTVVWVNTYNIIVTCHTPTFGGPKESGNG 491
QY 464 RELGEEGLKAYLEPKTINIHV 484
Db 492 RELGEDGLKAYTEVKTIVTIKV 512
RESULT 15
US-08-612-853-10
; Sequence 10, Application US/08612853
; GENERAL INFORMATION:
; APPLICANT: Achatz, Gernot
; APPLICANT: Oberkofler, Hannes
; APPLICANT: Simon, Birgit
; APPLICANT: Unger, Andrea
; APPLICANT: Lechenauer, Erich
; APPLICANT: Hirschwehr, Reinhold
; TITLE OF INVENTION: Recombinant Cladosporium Herbarum
; NUMBER OF SEQUENCES: 117
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/612,853
; FILING DATE: 26-FEB-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AT94/00120
; FILING DATE: 02-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Friebe, Thomas E
; REGISTRATION NUMBER: 29,258
; REFERENCE/DOCKET NUMBER: 6530-021-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 497 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Cladosporium herbarum
; US-08-612-853-10
Query Match 34.1%; Score 868; DB 10; Length 497;
Best Local Similarity 42.8%; Pred. No. 5.3e-81;
Matches 211; Conservative 67; Mismatches 183; Indels 32; Gaps 12;
QY 7 EYKSELFNNFVSSKGSERLTLPNWDSTVATDVHVANAADVDSAVAASVQAVKGPW 66
Db 15 EQPTGLFINNEFVKQEGKTFDVINPSDES-VITQVHEATEKDVDAVAAARQAF-EGSW 72
QY 67 KKFTGAQRAACMLKFADLAEKNAEKLARLESPLTGRPVSMITHFDIPNMVSV-----FRY 121

Db 73 RLETPENRGKLLNLANLANFEKNTDLLAAAVESLONGKATSM-----ARVTSACASGCLRY 126
QY 122 YAGWADKIAGKTFPEDNGKPNWRY---EPMGVACAGIASWNATFLYVGVWKIAPALAAAGCSF 178
Db 127 YGWMADKITGKVI--DTTPDTFNIVKKEPIGVCRSDHSLELPLLMWAWKIGPAIACGNTV 184
QY 179 IFKASEKSPGLVGLAPLFAEAGPPGWWQFLTG-ARVTGEALASHMDIAKISFTRSVGG 237
Db 185 VLKTAEQTPGLGGLVAASLYKEAGPPGVINVISGFGKVAGAAALSSHMDVDKVAFTGSTVV 244
QY 238 GRAVKQATLKSNNMKRVTLLEGER-PTIVFNEAPLERQ-SGESAKDPFSKFGQIWVPPSCLL 295
Db 245 GRTILKAAAASSNLKKVTLELGCKSPNIVFEDADIDNAISWVNFGIFFNHGQCCAGSRVY 304
QY 296 VQ---WGNLAEKFHGVRHGSFGGQORWLQGNPLEPKRTHGPFVDKSOYDRVLGNIDVGKD 352
Db 305 VQESIYDKFVQKFK-----ERQKNVVGDPFAADTFQGPQVSKVQFDRIMEYIQAGKD 357
QY 353 T-AQLLTGVGRKGDGFAIEPTIPVNPKPGSKIWFEIIFGPVLSIKTFKTEEEAIEIAND 411
Db 358 AGATVETGGSRKGDGYFIEPTIFSNVTEDMKIVKEEIFGPVCSIAKFKTKEDAIKLGA 417
QY 412 TTYGLASVIYTKSLNRGLRVSSALETGGVSINFPFIPETQTPFGGMKQSGGRELGEGL 471
Db 418 STYGLAAAAVHTKNLNTAIEVSNALKAGTVWVNTYHTLHHQMPFGGYKESGIGRELGEDAL 477
QY 472 KAYLEPKTINIHV 484
Db 478 ANYTQTKTVSIRL 490

Search completed: April 27, 2002, 07:55:20
Job time: 581 sec

QY 69 FTGAQRAACMLKPADLAEKNAEKLAERLESPT---GRPVSMITHFDIPNMVSVFRYYAGW 125
Db 96 MDASERGRLLYKLADLIERD-RLIATMESMESMGGLYSNAYLNDLAGCIKTLRYCAGW 154
QY 126 ADKI--AGKTFPED-NCKPNWRYEPGVCAGIASWNATLYVGWKIAPALAAACSFIFKA 182
Db 155 ADKIQQGRRTIPIDGNFFTYTRHEPIGVCQGIIIPWNFPLYMLIKIGPALSCGNTVVVKP 214
QY 183 SEKSPGLVIGLAPFAEAGFPVGVQFLTGARVT-GEALASHMDIAKISFTRSVGGGRAV 241
Db 215 AEOTPLTALHVASLIKEAGFPVGVVNIYPGYGPTAGAAISSHMDIDKVAFTGSTEVGKLI 274
QY 242 KQATLKSNNMKRVTLLELGEK-PTIVFNEAPLERQSGESAKDFSKFGQIWWPPSCLLVQWGN 300
Db 275 KEAAGKSNLKRVTLELGGKSPCIVLADADL-----DNAVEFAHHGVFYHQGCCIAASRI 329
QY 301 LAEK--FHGVRHGSFGGCQORWLGNQPNLEPKRTHGPFVDKSYDRVLGNIDVG-KDTAQLL 357
Db 330 FVEESIYDEFVRSVERAKKYIILGNPLTPGVTOGPOIDKEQYDKILDIESGKKEGAKLE 389
QY 358 TGVGRKGDGFAIEPTIFVNPKNPKSGKIWFEEIFGPVLSIKTFKTEEEAIEIANDTTYGLA 417
Db 390 CGGPGWNGKGYFQPTVFSNVTDEMRLAKEEIFGPVQIIMKFKSLDDVIKRANNTFYGLS 449
QY 418 SVIYTKSLNRLVRSSALETGGVSNFPFIPETQTPFGGMKQSGSGRELGEGLKAYLEP 477
Db 450 AGVFTKIDIKATISSALQAGTVVWVNCYGVVSAQCPCFGGFKMSGNGRELGEYGFHEYTEV 509
QY 478 KVINIHVN 485
Db 510 KTVTVKIS 517

RESULT 2

US-08-513-841-2
; Sequence 2, Application US/08513841
; Patent No. 5753481
; GENERAL INFORMATION:
; APPLICANT: Niwa, Mineo
; APPLICANT: Saito, Yoshimasa
; APPLICANT: Ishii, Yoshinori
; APPLICANT: Yoshida, Masaru
; APPLICANT: Suzuki, Hiromi
; TITLE OF INVENTION: No. 5753481el L-sorbose Dehydrogenase and No. 5753481el L-sorb
; TITLE OF INVENTION: Dehydrogenase Obtained from Gluconobacter oxydans T-100
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Oblon, Spivak, McClelland, Maier & Neustadt, P.C.
; STREET: 1755 Jefferson Davis Highway, Suite 400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MS-DOS Editor
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/513,841
; FILING DATE: 01-NOV-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: UK 9304700.9
; FILING DATE: 08-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 241851/1993
; FILING DATE: 28-SEP-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: NORMAN F. OBLON
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 18-909-0 PCT
; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; TELEX: 24855 OPAT UR
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 497 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Gluconobacter oxydans
; STRAIN: T-100
; FEATURE:
; NAME/KEY: mat peptide
; LOCATION: 1..497
; IDENTIFICATION METHOD: experimentally
; US-08-513-841-2

Query Match . 24.0%; Score 612.5; DB 1; Length 497;
Best Local Similarity 34.6%; Pred. No. 4.4e-60;
Matches 174; Conservative 83; Mismatches 197; Indels 49; Gaps 14;

QY 3 LSPDEYKSELFINNFRVSSKGSERLTLTNPWDESTVATDVHV-----ANAAVDOSAVAAAS 57
Db 11 LKPPREF--GFFIDGWRAGK-----DFDRSSPAHDVPVTRIPRCTREDLDEAVAAA 60
QY 58 VOAVKKGPKWKFTGAORAAACMLKFADLAEKNAEKLAERLESPTQRPVSMITHFOIPNMVS 117
Db 61 RRAFENGSWAGLAAADRAAVLLKAAGLLRRRDDIAYWEVLENGKPISQ-AKGEIDHCIA 119
QY 118 VFRYYAGWADKIAGKTFPEDNGKPNWRY-----EPMGVACAGIASWNATFLYVGWKIAPAL 172
Db 120 CFEMAAAGAARMHLRGDTF---NNLGEGLFGMVLRPEIGVVGLTPWNFPFMIICERAPFIL 176
QY 173 AAGCSFIFKASEKSPGLVGLAPLFAEAGFPVGVQFLTG-ARVTGEALASHMDIAKISF 231
Db 177 ASGCTLVVKPAEVTSATTLLEAILADAGLPKGVFNVTGTGRTVGOAMTEHQDIDMLSF 236
QY 232 TRSVGGGRAVKQATLKSNNMKRVTLLELGEK-PTIVFNEAPLERQSGESAKDFS-KFGQIWW 289
Db 237 TGSTGVGKSCIIHAAADSNNLKLGLLELGGKNPIVVFADSNLEDAADAVAFGISFNTGQCCV 296
QY 290 PPSCLLVQWGNLAEKF-----HGVRHGSFGGCQORWLGNQPNLEPKRTHGPFVDKSYD 341
Db 297 SSSRLIVE-RSVAERFERLVVPKMEKIRVG-----DPEDPETQIGAITTEAQNK 344
QY 342 RVLGNIDVGK-DTAQLLTGVGRKG-DKGFATIEPTIFVNPKNPKSGKIWFEEIFGPVLSIKTF 399
Db 345 TILDYIAKGAEGAKLLCGGGIVDFGKGQYIQTFLTVDVKPSMGIAARDEIFGPVLSAFHF 404
QY 400 KTEEEAIEIANDTTYGLASVIYTKSLNRLVRSSALETGGVSNFPFIPETQTPFGGMKQ 459
Db 405 DTVDEAIAIANDTVYGLAASVWSKIDIKALAVTRRVRAGRFRVNTIMSGGPETPLGGGFKQ 464
QY 460 SGSGRELGEGLKAYLEPKTINI 482
Db 465 SGWGREAGLYGVEEYTIKSVHI 487

RESULT 3

US-08-696-834-2
; Sequence 2, Application US/08696834
; Patent No. 5834263
; GENERAL INFORMATION:
; APPLICANT: Niwa, Mineo
; APPLICANT: Saito, Yoshimasa
; APPLICANT: Ishii, Yoshinori
; APPLICANT: Yoshida, Masaru
; APPLICANT: Hayashi, Hiromi
; TITLE OF INVENTION: Method for Producing 2-Keto-L-Gulonic Acid
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:


```

Query Match          24.0%; Score 612.5; DB 2; Length 497;
Best Local Similarity 34.6%; Pred. No. 4.4e-60;
Matches 174; Conservative 83; Mismatches 197; Indels 49; Gaps 14;

QY 3 LSPDEYKSELFINNEFVSSKGSERLTLTNPWDESIVATDVHV-----ANADVDSAVAAS 57
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 11 LKPREF--GFFIDGEWRACK-----DFFDRSPANDVPVTRIPRCTREDLDEAVAAA 60
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 58 VQAVKGPWKFTGAQRAACMLKFADLAEKNAEKLARLESPTGRPVSMITHFDIPNMVS 117
   : : : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 61 RRAFENGSWAGLAAADRAAVLLKAAGLLRERRDDIAYWEVLENGKPISQ-AKEIDHCIA 119
   : : : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 118 VERYYAGWADKIAGKTFEPDNGKPNWRY-----EPMGVACAGIASWNATFLVVGWKIAPAL 172
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 120 CFEMAAGAARMLHGDTF---NNLGEGLFGMVLREDPIGVGLITPWNFPEMILCERAPFIL 176
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 173 AAGSEIFKASEKSPLGVGLGLAPLFAEAGFPVGVQPLTG-ARVTGEALASHMDIAKISF 231
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 177 ASGCTLVVKPAEVTSATLLAEILADAGLPKGVFNVTGTGRTVVGQAMTEHQDIDMLSF 236
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 232 TRSVGGGRAVKQATLKSNNKRVTLLEGEK-PTIVNEAPLERQSGESAKDFS-KFGQIWW 289
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||
DB 237 TGSTGVGKSCIHAAADSNNLKLGLGELGKNPIVVFADSNLEDAADAVAFGISFNTGQCCV 296
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||

QY 290 PPSCLLVQWGNLAERF-----HGVRHGSFGGQORWLGNPLPKRTHGPFVDKSOYD 341
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||
DB 297 SSSRLIVE-RSVAERFERLVVPKMEKIRVG-----DPFDPETOICAITTEAQNK 344
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||

QY 342 RVLGNIDVGK-DTAQLLTGVGRKG-DKGFAIEPTIFVNPKPGSKIWFEEIFGPVLSIKTF 399
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||
DB 345 TILDYIAKGAEGAKLLCGGIVDFGKGQYIQIPLFTDVKPSMGIAERDEIFGPVLSAFHF 404
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||

QY 400 KTEERAIEIANDTTYGLASVIYTKSLNRLGLRVSSALETGGVSINFFPIPETQTPFGGMKQ 459
   ||| ||| ||| ||| ||| : : : ||| : : : ||| : : : ||| : : : ||| ||
DB 405 DTVDEAIAIANDTVYGLAASVWSKIDKALAVTRVRAGRFWVNTIMSGSGPPTPLGGFKQ 464
   ||| ||| ||| ||| ||| : : : ||| : : : ||| : : : ||| : : : ||| ||

QY 460 SSGRELGEGLKAYLEPKTINI 482
   ||| ||| ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||
DB 465 SGWCREAGLVGVEEYTIKSVHI 487
   ||| ||| ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||

RESULT 5
US-09-118-317-2
; Sequence 2, Application US/09118317
; Patent No. 6197562
; GENERAL INFORMATION:
; APPLICANT: Niwa, Mineo
; APPLICANT: Saito, Yoshimasa
; APPLICANT: Ishii, Yoshinori
; APPLICANT: Yoshida, Masaru
; APPLICANT: Suzuki, Hiromi
; TITLE OF INVENTION: No. 6197562el L-sorbose Dehydrogenase and No. 6197562el
; TITLE OF INVENTION: L-sorbose Dehydrogenase Obtained from Gluconobacter
; TITLE OF INVENTION: oxydants T-100
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Oblon, Spivak, McClelland, Maier & Neustadt, P.C.
; STREET: 1755 Jefferson Davis Highway, Suite 400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MS-DOS Editor
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/118,317
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/513,841

```

```

; FILING DATE: 01-NOV-1995
; APPLICATION NUMBER: UK 9304700.9
; FILING DATE: 08-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 241851/1993
; FILING DATE: 28-SEP-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: NORMAN F. ORLON
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 18-909-0 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; TELEX: 24855 OPAT UR
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 497 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Gluconobacter oxydans
; STRAIN: T-100
; FEATURE:
; NAME/KEY: mat peptide
; LOCATION: 1..497
; IDENTIFICATION METHOD: experimentally
;
US-09-118-317-2

Query Match 24.0%; Score 612.5; DB 4; Length 497;
Best Local Similarity 34.6%; Pred. No. 4.4e-60;
Matches 174; Conservative 83; Mismatches 197; Indels 49; Gaps 14;

Qy 3 ISPDEYKSELFINNEFYSSKGSERLTLTNPWDESTVATDVHV-----ANAAADVDSAAVAS 57
Db 11 LKPREF--GFFIDGWRAGK-----DFDRSSPAHDVPVTRIPRCTREDLDEAVAAA 60

Qy 58 VOAYKKGPKWKKFTGAORAAACMLKFADLAEKNAEKLARLES�PTGRPVSMITHEDIPNMVS 117
Db 61 RRAFENGSWAGLAAADRAAVLLKKAAGLLRRRDDIAYWEVLENGKPISO-AKGEIDHCIA 119

Qy 118 VFRYYAGWADKIAGKTFPEDNGKPNWRY-----EPMGVCAGIASWNATFLYVGWKIAPAL 172
Db 120 CFEMAAGAARMLHGDTF---NNLGEGLFGMVLREPIGVVGLITPHNPPFMILCERAPFIL 176

Qy 173 AACSFFIKASEKSPLCVLGLAPLFAEAGFPFGVVQFLTGT-ARVTGEALASHMDIAKISF 231
Db 177 ASGCTLVVKKPAEVTSATLLAEILADAGLPGKGVFNWVTGTGRTVQAMTEHQDIDMLSF 236

Qy 232 TRSYGGGGRVAKQATLKSNNKRVTTLEGEK-PTIVNEAPLERQSGESAKDFS-KFGQIYW 289
Db 237 TGSYGVGKSCIHAAADSNNLKLGLGKKNPIVVFADSNLEDAADAVAFGISFNTGQCCV 296

Qy 290 PPSCLLVQWGNLAEKF-----HGVRHGSFGGCQRWLQGNPLEPKRTHGPFVDKSOYD 341
Db 297 SSSRLIVE-RSVAEKFERLVVPKMEKIRVG-----OPFDPETOIGAITTEAQNK 344

Qy 342 RVLGNIDVGK-DTAQLLTGVGRKG-DKGFATEPTIFVNPKPGSKIWFEEIFGPVLSIKTF 399
Db 345 TILDYIAKGAEGAKLLCGGIVDFGKGQYIQPTLTFTDVKPSMGIGARDEIFGPVLASFHF 404

Qy 400 KTEEEAIEANDTTYGLASVIYTKSLNRLRVSSALETGGVSINFFPIPETQTPTFGGMKQ 459
Db 405 DTVEAIAIANDTVYGLAASVWSKDIKALAVTRRVAGRFRVWNTIMSGGPETPLGGFKQ 464

Qy 460 SSGRELGEGLKAYLEPKTINI 482
Db 465 SSGREAGLYGVEEYTOIKSVHI 487

```


OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/346,611
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Silvert, Donald J.
REGISTRATION NUMBER: 37552
REFERENCE/DOCKET NUMBER: 63632
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 616-5600
TELEFAX: (312) 616-5700
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 133 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-346-611-2

Query Match 9.1%; Score 231; DB 1; Length 133;
Best Local Similarity 43.7%; Pred. No. 5.le-18;
Matches 52; Conservative 19; Mismatches 46; Indels 2; Gaps 2;

QY 162 LYVGWKIAPALAAAGCSFIEKASEKSPGLVGLAPLFAEAGPPPGVVQFLTGARVT-GEAL 220
| ||: |||| | : | : ||: || | : || |||| | ||: || ||
Db 4 LMYAKVGPALACGNTLVKTAETPLSXLYISKLLHEAGLPEGVNVVSGFGPTAGAA 63
QY 221 ASHMDIAKISFTRSVGGGRAVKQATLKSNMKRVTLLEGEK-PTIVFNEAPLERQSGESA 278
||||: |||| | : | : ||: |||| | ||: || ||
Db 64 ASHMDVDKIAFTGSTDGTGKIILELAAKSNLKTVTLELGGKSPFFIIWTKLMFGPAXGAXA 122

RESULT 9
US-08-794-494-2
Sequence 2, Application US/08794494
Patent No. 5981833
GENERAL INFORMATION:
APPLICANT: Wise, Roger P.
APPLICANT: Schnable, Patrick S.
TITLE OF INVENTION: NUCLEAR RESTORER GENES FOR HYBRID SEED PRODUCTION
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Leydig, Voit & Mayer, Ltd.
STREET: Two Prudential Plaza, Suite 4900
CITY: Chicago
STATE: IL
COUNTRY: US
ZIP: 60601-6780
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/794,494
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 8-346611
FILING DATE: 29-NOV-1994
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Larcher, Carol
REGISTRATION NUMBER: 35243
REFERENCE/DOCKET NUMBER: 74959
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 616-5600
TELEFAX: (312) 616-5700
INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 133 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-794-494-2
Query Match 9.1%; Score 231; DB 2; Length 133;
Best Local Similarity 43.7%; Pred. No. 5.le-18;
Matches 52; Conservative 19; Mismatches 46; Indels 2; Gaps 2;

QY 162 LYVGWKIAPALAAAGCSFIEKASEKSPGLVGLAPLFAEAGPPPGVVQFLTGARVT-GEAL 220
| ||: |||| | : | : ||: || | : || |||| | ||: || ||
Db 4 LMYAKVGPALACGNTLVKTAETPLSXLYISKLLHEAGLPEGVNVVSGFGPTAGAA 63
QY 221 ASHMDIAKISFTRSVGGGRAVKQATLKSNMKRVTLLEGEK-PTIVFNEAPLERQSGESA 278
||||: |||| | : | : ||: |||| | ||: || ||
Db 64 ASHMDVDKIAFTGSTDGTGKIILELAAKSNLKTVTLELGGKSPFFIIWTKLMFGPAXGAXA 122

RESULT 10
US-07-689-008-2
Sequence 2, Application US/07689008
Patent No. 5268274
GENERAL INFORMATION:
APPLICANT: Ben-Bassat, Arie
APPLICANT: Calhoon, Roger D
APPLICANT: Fear, Anna L
APPLICANT: Gelfand, David H
APPLICANT: Meade, James H
APPLICANT: Tal, Rony
APPLICANT: Wong, Hing
APPLICANT: Ben-Ziman, Moshe
TITLE OF INVENTION: METHODS AND NUCLEIC ACID SEQUENCES FOR THE
TITLE OF INVENTION: EXPRESSION OF CELLULOSE SYNTHASE OPERON
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: McCutchen, Doyle, Brown & Enersen
STREET: Three Embarcadero Center
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/689,008
FILING DATE: 19910422
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 337,194
FILING DATE: 12-APR-1989
APPLICATION NUMBER:
APPLICATION NUMBER: US 496,236
FILING DATE: 23-MAR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Murphy, Lisabeth Feix
REGISTRATION NUMBER: 31547
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 393-2000
TELEFAX: (415) 393-2286
TELEX: 340817 MACPAG SFO
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 3031 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein

US-07-689-008-2

Query Match 3.7%; Score 94.5; DB 1; Length 3031;
Best Local Similarity 22.6%; Pred. No. 2.8;
Matches 67; Conservative 36; Mismatches 110; Indels 83; Gaps 17;
QY 198 AEAGFPVGVVQFLTGARTVGEALASHMDIAKISFTRSVGGRAVKQATLKSNNKRVTLLEL 257
Db 2440 AEAASPDTSQMLSDGGLGFRSRSGEHGMGRLT--- 2489
QY 258 GEKPTIVNEAPLERQSGESAKDFSKFGQIWPVPSCLLVQW-GNL-AEKFHGVHRHGSFG 314
Db 2490 -BANIPVIGRLPL-QAGASALFPS-----JTPDMI---WSGNLNTGSDVDPVRYGTMM 2537
QY 315 GQORWLGONPLEPKRTHGPFVDKSDYDRVLGNIDVGKDTAQLLGVGRKGDKGFAIEPTI 374
Db 2538 GVQAY-----NQYD---SYTNAGRDQORIAAGT---AEAGFA--PDV 2571
QY 375 -----FVNPKPGSK-----IWFEIFGVPVLSIKTFKTEEEAIEIAN----- 410
Db 2572 QFGNSWVRADVGASPIGFPITNVLGGEVFSRPGVPV----TFRVSAERRSITNSVLSYGG 2627
QY 411 -DTTYGLASVIYTKSL-NRGLRVSSALETGGVSVINEPFIPTOTPFPGMKQSGSG 463
Db 2628 LRDYNSALGRYARQVYQALSQWSEWGGVVYTNH-FHQOVEATLGNITLYGGG 2682

RESULT 11

US-08-220-958-4
; Sequence 4, Application US/08220958
; Patent No. 5459072
; GENERAL INFORMATION:
; APPLICANT: McKay, Larry
; APPLICANT: Polzin, Kayla
; TITLE OF INVENTION: FOOD-GRADE INTEGRATION VECTORS FOR
; INDUSTRIAL BACTERIAL STRAINS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MERCHANT & GOULD
; STREET: 3100 No. 5459072west Center
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/220,958
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/841,642
; FILING DATE: 25-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Kowalchuk, Alan W.
; REGISTRATION NUMBER: 31,535
; REFERENCE/DOCKET NUMBER: 600.229-US-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-332-5300
; TELEFAX: 612-332-9081
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 318 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Lactococcus lactis subsp lactis bv.

ORGANISM: diacetyllactis

STRAIN: DRC3
IMMEDIATE SOURCE:
CLONE: Putative Nisin Resistance Gene
US-08-220-958-4

Query Match 3.6%; Score 90.5; DB 1; Length 318;
Best Local Similarity 22.4%; Pred. No. 0.16;
Matches 55; Conservative 41; Mismatches 85; Indels 65; Gaps 14;
QY 225 DIAKISFTRSVGGRAVKQATLKSNNKRVTLLELGEKPTIVNEAPLERQSGESAKDFSKF 284
Db 104 DISKRSMTRYI-----KPKAEIEGNTLILTI-----PEFTGND-----QASDYANF 145
QY 285 QIWPVPSCLLVQWGNLAERKHGV---RHGSFGG---CQWLGONPLEPKRTHGPFVDKSO 339
Db 146 -----LESSLHKNNYNGVIVDLGRNCGDLSPLVGLSPLLPDGTLETYYVOKSS 194
QY 340 YDRVL-----GNIDVGKDTAQLLTVGVRKDKGFAIEP-TIFVNPKPGRK-----IWFEI 389
Db 195 HSKPVELQNGEINSGGSTKI-----SDNKKIKKAPIAVLIDNNTGSSGELTALCFEGI 248
QY 390 FGPVLSIKTFKTEEEAIEIANDTTYGLASVIYTKSLNRGLRVSSAL---ETGGVSVINFPF 446
Db 249 ---PNVKELGSDSAGYTSANQTVY-----LYDGST---LQITSFAVKDRTNNIYKNFPI 296
QY 447 IPETQT 452
Db 297 SPDIQT 302

RESULT 12

US-09-045-632-49
; Sequence 49, Application US/09045632
; Patent No. 6001575
; GENERAL INFORMATION:
; APPLICANT: Huanir, Richard L.
; APPLICANT: Dong, Hualing
; TITLE OF INVENTION: THERAPEUTIC USES OF GRIP AND
; TITLE OF INVENTION: GRIP-RELATED MOLECULES
; NUMBER OF SEQUENCES: 105
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
; STREET: 130 Water Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/045,632
; FILING DATE: 19-MAR-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/041,016
; FILING DATE: 19-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Corless, Peter F.
; REGISTRATION NUMBER: 33,860
; REFERENCE/DOCKET NUMBER: 48147/1699-CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-523-3400
; TELEFAX: 617-523-6440
; INFORMATION FOR SEQ ID NO: 49:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1050 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

MOLECULE TYPE: protein
US-09-045-632-49

Query Match 3.5%; Score 89.5; DB 3; Length 1050;
Best Local Similarity 20.1%; Pred. No. 1.6;
Matches 72; Conservative 47; Mismatches 102; Indels 137; Gaps 21;
QY 189 GVGLAPLFAEAGFPVGVQFL-----TGARVTGEALASHMDIAKISFTRSVGGGRA 240
DB 701 GPLGITISGTEEPDPIIISGLTKRGLAERTGASMLGTA---YWPSXSVSLK-----GRP 752
QY 241 VKQA--TLKSNMKRVTLLEGEKPTIVFNEAPLERQSG--ESAKDFSKFGQIWVPPSCL-- 294
DB 753 LSEAIHLQVAGETVTLKIKKQ---LDRPLLPRQSGSLSEASDVDE-----DPPEALKG 803
QY 295 -----LVQWGNLAEKFGHVGHSFGGC-----QRWLGQN 323
DB 804 GLLTTHFSPAVPSVDSAVESWGSSATE-----GGFGGSGSYTPQVAVRSVTPQEWSSR 857
QY 324 -----PLEPKRTH---GP-----FVDKSKQYD-----RVLGNI----- 347
DB 858 LKSSPPPLEPRRTSYTPGPTDESFPPEEEGDWEPPMSPAPGPAREEGFWRVLGEALEDLE 917
QY 348 DVGKD-----TAQLLTGVGRKGDKGFAIEPTIFVNPKPGSKIW-----FEEI 389
DB 918 SCGQSELLRELEASIMGTVQ-----SVAVDGRPGSRPWRRSREVCTSPEDLQEL 967
QY 390 FGPVLSIKTFKTEEEAIEIANDTTYGLASVIYTKSLN-RGLRVSSALETGGVSINFPF 446
DB 968 LLPT-PLEMHRVTLHKDPVRNDFGFSVSDGLEKGVYVHTVRIDGPAQHGGLO---PF 1021

RESULT 13

US-09-045-632-50
; Sequence 50, Application US/09045632
; Patent No. 6001575
; GENERAL INFORMATION:
; APPLICANT: Haganir, Richard L.
; APPLICANT: Dong, Hualing
; TITLE OF INVENTION: THERAPEUTIC USES OF GRIP AND
; NUMBER OF INVENTIONS: GRIP-RELATED MOLECULES
; NUMBER OF SEQUENCES: 105
; CORRESPONDENCE ADDRESS:
; ADDRESS: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
; STREET: 130 Water Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/045,632
; FILING DATE: 19-MAR-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/041,016
; FILING DATE: 19-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Corless, Peter F.
; REGISTRATION NUMBER: 33,860
; REFERENCE/DOCKET NUMBER: 48147/1699-CIP
; TELEPHONE: 617-523-3400
; TELEFAX: 617-523-6440
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1050 amino acids
; TYPE: amino acid

STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-045-632-50

Query Match 3.5%; Score 89.5; DB 3; Length 1050;
Best Local Similarity 20.1%; Pred. No. 1.6;
Matches 72; Conservative 47; Mismatches 102; Indels 137; Gaps 21;
QY 189 GVGLAPLFAEAGFPVGVQFL-----TGARVTGEALASHMDIAKISFTRSVGGGRA 240
DB 701 GPLGITISGTEEPDPIIISGLTKRGLAERTGASMLGTA---YWPSXSVSLK-----GRP 752
QY 241 VKQA--TLKSNMKRVTLLEGEKPTIVFNEAPLERQSG--ESAKDFSKFGQIWVPPSCL-- 294
DB 753 LSEAIHLQVAGETVTLKIKKQ---LDRPLLPRQSGSLSEASDVDE-----DPPEALKG 803
QY 295 -----LVQWGNLAEKFGHVGHSFGGC-----QRWLGQN 323
DB 804 GLLTTHFSPAVPSVDSAVESWGSSATE-----GGFGGSGSYTPQVAVRSVTPQEWSSR 857
QY 324 -----PLEPKRTH---GP-----FVDKSKQYD-----RVLGNI----- 347
DB 858 LKSSPPPLEPRRTSYTPGPTDESFPPEEEGDWEPPMSPAPGPAREEGFWRVLGEALEDLE 917
QY 348 DVGKD-----TAQLLTGVGRKGDKGFAIEPTIFVNPKPGSKIW-----FEEI 389
DB 918 SCGQSELLRELEASIMGTVQ-----SVAVDGRPGSRPWRRSREVCTSPEDLQEL 967
QY 390 FGPVLSIKTFKTEEEAIEIANDTTYGLASVIYTKSLN-RGLRVSSALETGGVSINFPF 446
DB 968 LLPT-PLEMHRVTLHKDPVRNDFGFSVSDGLEKGVYVHTVRIDGPAQHGGLO---PF 1021

RESULT 14

US-09-503-172A-2
; Sequence 2, Application US/09503172A
; Patent No. 6284510
; GENERAL INFORMATION:
; APPLICANT: ITO, Tetsuya
; APPLICANT: FUJITA, Koki
; APPLICANT: HARA, Kozo
; APPLICANT: TONOSUKA, Takashi
; APPLICANT: SAKANO, Yoshiyuki
; TITLE OF INVENTION: BETA-FRUCTOFURANOSIDASE GENE
; FILE REFERENCE: 10749-0001-0
; CURRENT APPLICATION NUMBER: US/09/503,172A
; PRIOR FILING DATE: 2000-02-14
; PRIOR APPLICATION NUMBER: JP 160416/1999
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 2
; LENGTH: 578
; TYPE: PRT
; ORGANISM: Arthrobacter sp.
US-09-503-172A-2

Query Match 3.5%; Score 88.5; DB 4; Length 578;
Best Local Similarity 22.2%; Pred. No. 0.76;
Matches 121; Conservative 49; Mismatches 193; Indels 181; Gaps 28;
QY 20 SSKGSERLTNPWDESTVATDVHANAADVDSAVAASVQAVKKGPKWKFTGAQRAACML 79
DB 4 STRGRVVRVLAGGLATSTLAAAVLIAGAAP-----ATAQSGLDGDPETIHTQQAYAPED 58
QY 80 KF-ADLAEKNAEKLARLESPT--GRPVSMITHFDIPNMVSF----- 119
DB 59 DFTAKWTRADARQLORM-SDPTAPSRNSMPASVTMTVPQDFPDMSNEQVWVWDTWPLT 117
QY 120 ----RYYA--GW-----ADKIAGKTFPEDNCKPNWRYEPMGVCAIASWNATFLYVGW 166

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2002, 07:45:31 ; Search time 104.96 Seconds
(without alignments)
515.367 Million cell updates/sec

Title: US-09-882-694A-5
Perfect score: 2549
Sequence: 1 MVLSPDEYKSELFINFVFS.....EEGLKAYLEPKTINHVNIE 487

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802.*
1: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2549	100.0	487	22	AA72635	Exophiala spinifer
2	866	34.0	496	16	AA71891	Cladosporium herba
3	857	33.6	497	7	AA660456	Sequence of aldehy
4	824.5	32.3	529	22	ABG06577	Novel human diagno
5	818.5	32.1	500	22	AA74924	Human aldehyde deh
6	817.5	32.1	521	15	AA63672	Aldehyde-dehydroge
7	815.5	32.0	495	16	AA71803	Alternaria alterna
8	781.5	30.7	501	21	AAG36239	Arabidopsis thalia
9	781.5	30.7	501	21	AA767413	Arabidopsis aldehy
10	781	30.6	538	21	AA767412	Arabidopsis aldehy
11	776.5	30.5	520	22	ABB60140	Drosophila melanog

12	772.5	30.3	659	22	ABB62256	Drosophila melanog
13	725	28.4	534	21	AA767414	Arabidopsis aldehy
14	704	27.6	909	22	ABB58519	Drosophila melanog
15	700	27.5	424	21	AAG36240	Arabidopsis thalia
16	684	26.8	495	22	AA74923	Yeast aldehyde deh
17	681.5	26.7	497	21	AAG16882	Arabidopsis thalia
18	681.5	26.7	501	21	AAG16881	Arabidopsis thalia
19	666.5	26.1	496	22	AAG81968	S. epidermidis ope
20	665	26.1	490	22	AAU34464	E. coli cellular p
21	660.5	25.9	496	22	AAU34148	Staphylococcus aur
22	660.5	25.9	496	22	AAU37064	Staphylococcus aur
23	655.5	25.7	503	21	AAG35662	Arabidopsis thalia
24	655.5	25.7	503	21	AAG52517	Arabidopsis thalia
25	655.5	25.7	532	21	AAG35661	Pseudomonas aerugi
26	645	25.3	490	22	AAU36509	Micromonospora eve
27	629.5	24.7	477	22	AAU04880	Staphylococcus aur
28	619.5	24.3	508	18	AAW89793	L-sorbose dehydr
29	612.5	24.0	497	15	AAR63766	Barley betaine ald
30	612.5	24.0	505	17	AAR97731	Salmonella typhi c
31	606.5	23.8	481	22	AAU38236	Aldehyde dehydrogen
32	605.5	23.8	508	21	AA802082	Pseudomonas putida
33	604.5	23.7	507	22	AAG80035	Pseudomonas putida
34	604.5	23.7	507	22	AAG80043	Gluconobacter oxyd
35	601.5	23.6	497	16	AAR82542	R. ruber 12-oxo la
36	599.5	23.5	474	22	AAB85329	Arabidopsis thalia
37	587	23.0	357	21	AAG36241	C glutamicum prote
38	582.5	22.9	497	22	AAG90354	Corynebacterium gl
39	582.5	22.9	497	22	AA791113	E. coli aldehyde d
40	578.5	22.7	512	22	AA74926	Enterococcus faeca
41	575	22.6	496	22	AAU34957	Arabidopsis thalia
42	572	22.4	391	21	AAG16883	Human prostate can
43	569	22.3	314	21	AAB56770	Arabidopsis thalia
44	566	22.2	393	21	AAG35663	Arabidopsis thalia
45	566	22.2	393	21	AAG52518	Arabidopsis thalia

ALIGNMENTS

RESULT 1
AA72635
ID AA72635 standard; Protein; 487 AA.
XX
AC AA72635;
XX
DT 02-MAY-2001 (first entry)
XX
DE Exophiala spinifera aldehyde dehydrogenase.

XX Fumonisin; degradative enzyme; transport enzyme; flavin monooxygenase;
KW aldehyde dehydrogenase; permease; p-glycoprotein; fumonisin esterase;
KW detoxification; mycotoxin; animal feed; human feed; silage;
KW transgenic plant; transgenic animal; microbial spray.

OS Exophiala spinifera.

XX WO200105980-A1.

PD 25-JAN-2001.

XX 14-JUL-1999; 99WO-US15824.

XX 14-JUL-1999; 99WO-US15824.

XX (PION-) PIONEER HI-BRED INT INC.

XX (CURA-) CURAGEN CORP.

XX Duvick JP, Maddox J, Gilliam J, Folkerts O, Crasta OR;

XX WPI; 2001-147345/15.

XX N-PSDB; AAD02692.

XX Novel polynucleotides encoding Exophiala degradative or transport

PT enzyme which is useful for detoxifying fumonisin or structurally
PT related mycotoxin during processing of grain for human or animal food
XX consumption -
PS Claim 1c; Page 68-70; 90pp; English.
XX
CC The patent discloses novel polynucleotides encoding Exophiala spinifera
CC fumonisin degradative or transport enzymes such as flavin monooxygenase,
CC aldehyde dehydrogenase, permease and p-glycoprotein. A fumonisin esterase
CC enzyme with at least one fumonisin degradative enzyme is useful for
CC detoxifying fumonisin or a structurally related mycotoxin present in
CC harvested grain, on application to a plant or to harvested grain
CC during processing, or to processed grain that is to be used as animal
CC or human feed, or as a silage. Nucleotide fragments of the present
CC invention are useful as probes and primers. They can be introduced
CC into microorganisms that multiply on plants to deliver enzymes to
CC potential target crops. The genes encoding the degrading enzymes are
CC introduced via a vector into a microbial host and the transformed host
CC is supplied to the environment, plants or animals for reducing the
CC pathogenicity of a fungus producing fumonisin. The genes of the
CC invention are fermented in a bacterial host and the resulting bacteria
CC is processed and used as a microbial spray. The nucleotide sequences
CC can be used alone or in combination to engineer microbes or other
CC organisms to metabolise fumonisin and resist its toxic effects.
CC The present protein sequence is aldehyde dehydrogenase, a fumonisin
CC degradative enzyme from Exophiala spinifera.
XX
SQ Sequence 487 AA;

Query Match 100.0%; Score 2549; DB 22; Length 487;
Best Local Similarity 100.0%; Pred. No. 2.5e-237;
Matches 487; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVLSPDEYKSELFINNFEVSSKGSERLTLTNPWDESTVATDVHVANAADVDSAVAASVQA 60
Db 1 mvlspdeykselfinnfevsskgserltnpwestvatdvhvanaadvdsavaasvqa 60
QY 61 VKKGPWKKFTGAQRAACMLKPADLAEKNAEKLARLESPLTGRPVSMITHFDIPNMVSFVR 120
Db 61 vkkgpwkkftgaqraacmlkpadlaeknaeklarlespltpgrpvsmithfdipnmvsfvr 120
QY 121 YYAGWADKIAGKTFPEDNGKPNWRYEPMGVCAIASWNATFLYVGWKIAPALAAAGSPFIF 180
Db 121 yyagwadkiagktrpedngkpnwryepmgvcagiaswnatflyvgwkiapalaaagsfif 180
QY 181 KASEKSPGLVGLAPLFAEAGFPPGVVOFLTGARTGEALASHMDIAKISFTRSVGGGRA 240
Db 181 kasekspglvlgaplfaeagfppgvvqltgartvgealashmdiakisftrsvgggra 240
QY 241 VKQATLKSNNMKRVLTLELGEKPTIVNEAPLERQGESAKDFSKFGQIWVPPSCLLVQWGN 300
Db 241 vkqatlksnmkrvltlelgekptivneaplerqgesakdfskfgqiwpvppscillvqwn 300
QY 301 LAEKFHGVRHSGFCGQCQRLWLGONPLEPKRTHGPFVDKSOYDRVLGNIDVGKDTAQLLTGV 360
Db 301 laekfhgvrhsgfcgqcqrlwlgonplepktrthgpfvdksqydrvlgnidvgkdtqlltgv 360
QY 361 GRKGDKGFAIEPTIFVNPKPGSKIWFEEIFGFPVLSIKTFKTEEEAIEIANDTTYGLASVI 420
Db 361 grkgdkgfaieptifvnpkpgskiwfeefgfpvlsiktfkteeeaielandttyglasvi 420
QY 421 YTKSLNRLGRVSSALETGGVSINFPFIPETQTPFGGMKQSGSGRELGEGLKAYLEPKTI 480
Db 421 ytkslnrglrvssaletggvsinfpfpetqtpfggmkgsgsgrelgeeglkaylepkti 480
QY 481 NIHVNIE 487
Db 481 nihvnie 487

RESULT 2
AAR71891

ID AAR71891 standard; Protein; 496 AA.
XX AAR71891;
AC AAR71891;
DT 25-OCT-1995 (first entry)
XX Cladosporium herbarum allergen Clah53.
DE Fungal spore; allergen; Clah53; allergy; aldehyde dehydrogenase.
KW Cladosporium herbarum.
OS WO9506121-A.
XX PN 02-MAR-1995.
PD 24-AUG-1994; 94WO-AT00120.
XX PF 27-AUG-1993; 93AT-0001725.
PR (BIOM-) BIOMAY PRODN & HANDELSGES MBH.
XX PA Achatz G, Breitenbach M, Ebner C, Hirschwehr R;
PI Kraft D, Lechenauer E, Oberkofler H, Prillinger H;
PI Simon B, Unger A;
XX DR WPI; 1995-106850/14.
DR N-PSDB; AAQ86278.
XX Allergens derived from Cladosporium herbarum spores - also
PT recombinant DNA for expressing the allergens, useful for in vitro
PT allergy detection
XX Claim 1; Page 8-9; 35pp; German.
PS Spores of Cladosporium herbarum are the most common fungal spores
XX found in the air; they can cause allergic reactions. Various Clah
CC allergens and sequences encoding them have now been isolated. The
CC mature Clah53 allergen has mol. wt. 53 kD and is encoded by cDNA
CC sequence AAQ86278. The allergen has homology to aldehyde
CC dehydrogenases. Potential epitopic subfragments were identified by
CC computer analysis of the amino acid sequence. See AAR71892-R71906 for
CC potential B-cell epitopes and AAR72615-R72627 for potential T-cell
CC epitopes.
XX SQ Sequence 496 AA;
Query Match 34.0%; Score 866; DB 16; Length 496;
Best Local Similarity 42.8%; Pred. No. 9.7e-75;
Matches 211; Conservative 67; Mismatches 183; Indels 32; Gaps 12;
QY 7 EYKSELFINNEFVSSKGSERLTLTNPWDESTVATDVHVANAADVDSAVAASVQAVKKGWP 66
Db 15 eqptglfinnefvskgqegktfdvinpsdes-vitqvheatekdvdlaavaargaf-egsw 72
QY 67 KKFTGAQRAACMLKPADLAEKNAEKLARLESPLTGRPVSMITHFDIPNMVSV-----FRY 121
Db 73 rletpenrgkllnnlanlfekntdliaavesldngkatsm-----arvtsacasgclry 126
QY 122 YAGWADKIAGKTFPEDNGKPNWRY---EPMGVCAIASWNATFLYVGWKIAPALAAAGCSF 178
Db 127 yggwadkitgkvi--dtptdfnyvkkepgvcrsdhslelpllmwawkigpalacntv 184
QY 179 IFKASEKSPGLVGLAPLFAEAGFPPGVVOFLTG-ARVTGEALASHMDIAKISFTRSVGG 237
Db 185 vlktaeqtplgglvaaslvkeagfppgvvinvisgfgkvagaalsshmdvdkvftgstvv 244
QY 238 GRAVKQATLKSNNMKRVLTLELGEK-PTIVNEAPLERQ-SGESAKDFSKFGQIWVPPSCLL 295
Db 245 grtilkaaassnlkkvtlelglgkspnlvifedadidnaiswnfglffnhgqccagrvy 304
QY 296 VQ---WGNLAEKFHGVRHSGFCGQCQRLWLGONPLEPKRTHGPFVDKSOYDRVLGNIDVGKD 352

[illegible]

QY	72	AQRAACMLKPADLAENAEKNAEKLARLESLSPTGRPSVMITHFDIPNMVSVFRYYAGWADKIAG	131
Db	79	sergilinkiadlmerdidtlaaiesldngkaftm-akvdiansigclryyagwackihg	137
QY	132	KTFPEDNGKPNW-RYEPMGVCAGIASWNATFLYVGWKIAPALAAAGCSFIFKASEKSPGLGV	190
Db	138	qtidtnpetitytrhepvvgcgqiipwnfpllmwswkigpavaagntvvlktaqgtplsa	197
QY	191	LGLAPLFAEAGFPVGVQFLTG-ARVTGEALASHMDIAKISFTRSVGGGRAVKQATLKSN	249
Db	198	lyaaaklikeappagvinvisgfgtagaaisshmdidkvaftgstlvgtptilqaaaksn	257
QY	250	MKRVTLLEGK-PTIVFNEAPLERQ-SGESAKDFSKEGQIWVPPSCLLVQWGNLAEKFP--	305
Db	258	lkkvtlelggkspnivfddadidhaishwanfgiffnhggqccagrsrlvqeg-iydxkfa	316
QY	306	--HGVRHGSFGGCGQRWLGNQNPLEPRXTHGPFVDKSOYDRVLGNIDVGKDT-AQLLTGVG	361
Db	317	rfkeraqknkvg-----npfeqdtfggpqvsqldrimeyinhgkagatvatcgdd	368
QY	362	RKGDKGFAIETPIFVNPKPGSKIWFEIEIFGPVLSIKTFKTEBAAIEIANDTTYGLASVIY	421
Db	369	rhgneyfiqptvftdvtsdmkiageeifgpvvtiqlkdvaaikigxstdyglaaavh	428
QY	422	TKSLNRGLRVSSALETGGVSINFPFIPDETQTFFGGMKQSGSRELGEGLKAYLEPKTIN	481
Db	429	tknvtairysnalkagtvinnynmisyqapfggfkqsgigrelgsyalenytqiktvh	488
RESULT 4			
QY	ABG06577	standard; Protein; 529 AA.	
XX	AC	ABG06577;	
XX	DT	13-FEB-2002 (first entry)	
XX	DE	Novel human diagnostic protein #6568.	
XX	KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;	
XX	KW	food supplement; medical imaging; diagnostic; genetic disorder.	
XX	OS	Homo sapiens.	
XX	PN	WO200175067-A2.	
XX	PD	11-OCT-2001.	
XX	PF	30-MAR-2001; 2001WO-US08631.	
XX	PR	31-MAR-2000; 2000US-0540217.	
XX	PR	23-AUG-2000; 2000US-0649157.	
XX	PA	(HYSE-) HYSEQ INC.	
XX	PI	Drmanac RT, Liu C, Tang YT;	
XX	DR	WPI; 2001-639362/73.	
XX	DR	N-PSDB; AAS70764.	
XX	PT	New isolated polynucleotide and encoded polypeptides, useful in	
XX	PT	diagnostics, forensics, gene mapping, identification of mutations	
XX	PT	responsible for genetic disorders or other traits and to assess	
XX	PT	biodiversity -	
XX	PS	Claim 20; SEQ ID No 36936; 103pp; English.	
XX	CC	The invention relates to isolated polynucleotide (I) and	
XX	CC	polypeptide (II) sequences. (I) is useful as hybridisation probes,	
XX	CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosom	
XX	CC	and gene mapping, and in recombinant production of (II). The	
XX	CC	polynucleotides are also used in diagnostics as expressed sequence ta	
XX	CC	for identifying expressed genes. (I) is useful in gene therapy techni	

Qy 298 WGNLAEKFHGVRHGSFGGQCRWLGNQNLPEPKRTHGPFVDKSYQDRVLGNIDVGK-DTAQL 356
Db 313 -ediydef--vvrsvaraksvvg-npfdskteggpvdetqfkkilgyintgkqegakl 368
Qy 357 LTGVRKGDKGFAIEPTIFVNPKPGSKIWFEEIFGPVLSIKTFKTEEEAIEIANDTTYGL 416
Db 369 lcggglaadryfqlptvfgdvqdgmtlakeelfgpmqlkfkteevvgrannstyg1 428
Qy 417 ASVIYTKSLNRGLRVSSALETGGVSINFPFIPETQTPFGGMKQSGSGRELGEGLKAYLE 476
Db 429 aaavftkdldkanylsqalqagtvwvncydvfgaqspfggykmsgsgrelgeyglqayte 488
Qy 477 PKTINIHV 484
Db 489 vktvtkv 496
RESULT 6
AAR63672
ID AAR63672 standard; Protein; 521 AA.
XX
AC AAR63672;
XX
DT 08-MAY-1995 (first entry)
XX
DE Aldehyde-dehydrogenase Aldh-1.
XX
KW Retro virus; vector; aldehyde-dehydrogenase;
KW glutamylcysteine-synthetase; hematopoietic cell; cyclophosphamide;
KW chemotherapy; transgenic animal; gene therapy; cancer therapy;
KW selectable marker.
XX
OS Homo sapiens.
XX
PN WO9423015-A.
XX
PD 13-OCT-1994.
XX
PF 01-APR-1994; 94WO-US03624.
XX
PR 01-APR-1993; 93US-0041722.
XX
PA (UYCO) UNIV COLUMBIA NEW YORK.
XX
PI Dalla-favera R, Gianni AM;
XX
DR WPI; 1994-333177/41.
DR N-PSDB; AAQ72450.
XX
PT Retroviral vectors encoding human cytosolic aldehyde
PT dehydrogenase or glutamyl cysteine synthetase - used to transform
PT a subject's haematopoietic cells to reduce the toxic effects of
PT cyclo phosphamide chemotherapy
XX
PS Disclosure; Fig.4; 92pp; English.
XX
CC A novel retro virus vector encodes human cytosolic aldehyde-
CC dehydrogenase and/or human glutamylcysteine-synthetase (AAR63673).
CC Hematopoietic cells transfected by the vector are resistant to
CC cyclophosphamide, providing a means of gene therapy that allows
CC higher doses of toxic drugs to be used in cancer chemotherapy.
CC The human genes may also be used as selectable markers for
CC mammalian cell transfection and for transgenic animal breeding.
XX
SQ Sequence 521 AA;
Query Match 32.1%; Score 817.5; DB 15; Length 521;
Best Local Similarity 39.3%; Pred. No. 5e-70;
Matches 192; Conservative 90; Mismatches 187; Indels 19; Gaps 11;
Qy 10 SELFINNEFVSSKGSERLFLTNPWDESTVATDVHVANAADVDSAVAASVQAVKKG-PWKK 68
:::|||||: | :: :|| | : | : ||| || | : | | :

Db 37 tkifinnewhdsvsgkfkpfvfnpateeele-qveegdkedvdkaavqafqigspwrt 95
Qy 69 FTGAQRAACMLKFADLAEKNAEKLARLES1PT---GRPVSMITHFDIPNMVSFVRYVAGW 125
Db 96 mdasergrllykiallierd-rllatmesmesmnggklysnayindlagciktlyrcagw 154
Qy 126 ADKI--AGKTFPED-NGKPNWRYEPMGVCAGIASWNATFLYVGWKIAPALAAAGCSFIFKA 182
Db 155 adkiqqgrrtipidgnfftytrhepivgcqilpwnfplvmlwiikgpalscgntvvvvp 214
Qy 183 SEKSPLGVGLAPLFAEAGFPVGVVQFLTGARVT-GEALASHMDIAKISFTRSVGGGRAV 241
Db 215 eeqtpltalhvaslikeagfppgvvniwpgygptagaaissshmdidkvaftgstevgkii 274
Qy 242 KOATLKSNMKRVTLLEIGEK-PTIVFNEAPLFRQSGESAKDFSKFGQIWWPPSCLLVQWGN 300
Db 275 keaagksnlkrvtielggkspcivladadl-----dnavefahghvfyhqgqcciaasri 329
Qy 301 LAEK--FHGVRHGSFGGQCRWLGNQNLPEPKRTHGPFVDKSYQDRVLGNIDVG-KDTAQLL 357
Db 330 fveesiydefvrrsverakkyilgnpltpgvttgppqidkeqydkildliesgkkegakle 389
Qy 358 TGVGRKGDKGFAIEPTIFVNPKPGSKIWFEEIFGPVLSIKTFKTEEEAIEIANDTTYGLA 417
Db 390 cgggpgwnkgyfvqtvfsvnvtmdemriakeelfgpvqqimkfsiddvikranntfygis 449
Qy 418 SVIYTKSLNRGLRVSSALETGGVSINFPFIPETQTPFGGMKQSGSGRELGEGLKAYLEP 477
Db 450 agvftkdldkaltissalqagtvwvncygvvsaqcpfggfkmsgngrelgeygfheytev 509
Qy 478 KTINIHV 485
Db 510 kvtvtkis 517
RESULT 7
AAR71803
ID AAR71803 standard; Protein; 495 AA.
XX
AC AAR71803;
XX
DT 02-NOV-1995 (first entry)
XX
DE Alternaria alternata allergen Alta53.
XX
KW Fungal spore; allergen; Alta53; allergy; aldehyde dehydrogenase.
XX
OS Alternaria alternata.
XX
PN WO9506122-A.
XX
PD 02-MAR-1995.
XX
PF 24-AUG-1994; 94WO-AT00121.
XX
PR 27-AUG-1993; 93AT-0001726.
XX
PA (BIOM-) BIOMAY PRODN & HANDELSGES MBH.
XX
PI Achatz G, Breitenbach M, Ebner C, Hirschwehr R;
PI Kraft D, Lechenauer E, Oberkofler H, Prillinger H;
PI Simon B, Unger A;
XX
DR WPI; 1995-106851/14.
DR N-PSDB; AAQ86275.
XX
PT Allergens derived from Alternaria alternata - their isolation by
PT reacting sera from patients with fungal extracts; useful for
PT in-vitro allergy detection.
XX
PS Claim 7; Page 8; 26pp; German.
XX
CC Spores of Alternaria alternata can cause allergic reactions. Various

XX The invention provides nucleic acids encoding Arabidopsis plastidic
CC acetyl CoA synthetase (ACS), various subunits (specifically the E3
CC subunit) of plant plastidic pyruvate dehydrogenase (ppdh), the A and B
CC subunits of a plant ATP citrate lyase (ACL), Arabidopsis pyruvate
CC decarboxylase (PDC), Arabidopsis aldehyde dehydrogenase (ALDH),
CC specifically ALDH-2 and ALDH-4. The polypeptides can be expressed by
CC standard recombinant methodology. The ACS, ppdh, ACL, PDC and ALDH
CC polypeptides, methods and nucleic acid molecules of the invention are
CC used to alter the level of acetyl CoA in a plant or plant cell, tissues
CC or organs. A decrease in acetyl CoA is expected to affect the
CC biosynthesis of very long chain fatty acids and flavonoids. The enzymes
CC may also be used for in vitro synthesis of acetyl CoA, which in turn can
CC be used to produce acetyl CoA phytochemicals (plastidic ACS, ppdh, ACL,
CC pyruvate decarboxylase, acetyl CoA hydrolase, mitochondrial pyruvate
CC dehydrogenase and aldehyde dehydrogenase are involved in the biosynthesis
CC of acetyl CoA.
XX
SQ Sequence 538 AA;

Query Match 30.6%; Score 781; DB 21; Length 538;
Best Local Similarity 38.6%; Pred. No. 1.8e-66;
Matches 188; Conservative 81; Mismatches 196; Indels 22; Gaps 11;

QY 10 SELFVSSKSGSERLTLTNPWDESTVATDVHVA--NAADVSAVAASVQAVKGPWK 67
Db 58 tqllingnfvdasgkftptldprtgevia--hvaegdaedlnravkaartaafdegwp 114
QY 68 KFTGAQRAACMLKFADLAEKNAEKLARLESPTGRPVSMITHFDIPNMVSFVRYAGWAD 127
Db 115 kmsayersrvllrfadlvkehseelasletwngkpyqqsldaelpmfarlfryyagwad 174
QY 128 KIAGKTFPED-NGKPNWRYEPMGVCGACIASWNATFLYVGWKIAPALAAAGCSFIFKASEKS 186
Db 175 kimgltipadgnyqvtlhhepivagqilipwnfpllmfawkvgalacgntivlktaeqt 234
QY 187 PLGVLGLAPLFAEAGPPPGVVQFLTGTARVT-GEALASHMDIAKISFTRSVGGGRAVKQAT 245
Db 235 pltafyagklfleaglpvgvlnivsgfgatagaalashmdvdklaftgtdtgkvilgla 294
QY 246 LKSNMKRVLTLELGEK-PTIVFNEAPLERQSGESAKDFSKFGQIWVPPSCLLVQWGN---L 301
Db 295 ansnlkpvtlelgkspfvifedadidk----avelahfalffnqggqccca--gsrtfv 347
QY 302 AEKFHG--VRHSGFGGQCRWLGNQNPLEPKRTHGPFVDKSOYDRVLGNIDVG-KDTAQLLT 358
Db 348 hekvydefvekskaralkrvvg-dpfrkgieggpqiidlqfekvmkyiksgiesnatiec 406
QY 359 GVKRKGDKGFAIEPTIFVNPKPCKSGKIWFEEIFGFPVLSEIKTFKTEEEAIEIANDTTYGLAS 418
Db 407 ggdqigdkgyfiqptvfnsvkddmliaqdelgfpvqsilksfsvdevikranetkyglaa 466
QY 419 VIYTKSLNRGLRVSSALETGGVSINFFPIPETOTPPGGMKQSGSGRELGEGLKAYLEPK 478
Db 467 gvftknldtanrvsralkagtvvwncfdvdaaipfggymsgngrekgiyslnnylqix 526
QY 479 TINIHVN 485
Db 527 avvtaln 533

RESULT 11
ABB60140
ID ABB60140 standard; Protein; 520 AA.
XX
AC ABB60140;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 7212.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL04243.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 7212; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 520 AA;

Query Match 30.5%; Score 776.5; DB 22; Length 520;
Best Local Similarity 37.5%; Pred. No. 4.6e-66;
Matches 183; Conservative 91; Mismatches 201; Indels 13; Gaps 11;

QY 4 SPDEYKSELFVSSKSGSERLTLTNPWDESTVATDVHVAADVDSDAVASVQAVKK 63
Db 34 tpdilytgvfinnwhksksgkifetinpttaevia-eiqcadkedidlavqaarnaflk 92
QY 64 G-PWKKFTGAQRAACMLKFADLAEKNAEKLARLESPTGRPVSMITHFDIPNMVSFVRY 122
Db 93 gspwrrmdasergrllyrladlmerdqvyilasletldngkpsysynvndlptaiknlryf 152
QY 123 AGWADKIAGKTFPEDNGKPNW-RYEPMGVCAGIASWNATFLYVGWKIAPALAAAGCSFIFK 181
Db 153 agwadknhgktipmdgdfftytrhepvvgvqilipwnfpillmmawklgpalatgntivlk 212
QY 182 ASEKSPJGVGLGLAPLFAEAGPPPGVVQFLTGTARVTGEALASHMDIAKISFTRSVGGGRAV 241
Db 213 paeqtsltalyiaqlvkeagfpegvvnvvpvgfgtagaalanhcdvkvafstgdvgkli 272
QY 242 KQATLKSNMKRVLTLELGEK-PTIVFNEAPLERQSGESAKD--FSKFGQIWVPPSCLLVQW 298
Db 273 qlasgntnlkrvtielgkspnillsdtdmd-yavetahfglfnmgqcccgartfve- 330
QY 299 GNLAEKFHGVRHSGFGGQCRWLGNQNPLEPKRTHGPFVDKSOYDRVLGNIDVG-KDTAQLL 357
Db 331 dkiydef--versaerakkrvtg-npfdInteggpqvneeqmekilgmiktgkqgaklv 387
QY 358 TGVGR-KGDKGFAIEPTIFVNPKPCKSGKIWFEEIFGFPVLSEIKTFKTEEEAIEIANDTTYGL 416
Db 388 aggsrpeglpgyfvqptvfadvqdmdmtiareelfgvpvqqlirfxkidevierannseygl 447

CC subunits of a plant ATP citrate lyase (ACL), Arabidopsis pyruvate
CC decarboxylase (PDC), Arabidopsis aldehyde dehydrogenase (ALDH),
CC specifically ALDH-2 and ALDH-4. The polypeptides can be expressed by
CC standard recombinant methodology. The ACS, PPDH, ACL, PDC and ALDH
CC polypeptides, methods and nucleic acid molecules of the invention are
CC used to alter the level of acetyl CoA in a plant or plant cell, tissues
CC or organs. A decrease in acetyl CoA is expected to affect the
CC biosynthesis of very long chain fatty acids and flavonoids. The enzymes
CC may also be used for in vitro synthesis of acetyl CoA, which in turn can
CC be used to produce acetyl CoA phytochemicals (plastidic ACS, PPDH, ACL,
CC pyruvate decarboxylase, acetyl CoA hydrolase, mitochondrial pyruvate
CC dehydrogenase and aldehyde dehydrogenase are involved in the biosynthesis
CC of acetyl CoA.
XX
SQ Sequence 534 AA;

Query Match 28.4%; Score 725; DB 21; Length 534;
Best Local Similarity 37.2%; Pred. No. 4.5e-61;
Matches 181; Conservative 82; Mismatches 193; Indels 30; Gaps 12;

QY 10 SELFINNEFVSSKGSERLTLTNPWDESTVATDVHVANAADVDSAVAASVQAVKKGPKWKF 69
Db 54 tqlllggrfvdavsgktftldprngevia-qvsegdaedvnrvavaaarkafdegwpk 112
QY 70 TGAQRAACMLKFADLAEKNAEKLARLESPTGRPPVSMITHFDIPNMVSVFRYYAGWADKI 129
Db 113 tayerskilfrfadliekhnndelaaletdwngkpyeqsaqievmlarfryyagwadki 172
QY 130 AGKTFPEDNGKPN--WRVEPMGVCAGIASWNATFLYGVGKFIAPALAACGSFIFKASEKS 186
Db 173 hgmtmpgdg--phhvqtlhepigvagqilpwnfplmlswklgpalachgntvvlktaeqt 230
QY 187 PLGVLGLAPLFAEAGFPFGVQFLTGARVT-GEALASHMDIAKISFTFRSVGGGRAVKQAT 245
Db 231 plsalivgkllheaglpdgvvnivsgfgatagaalashmdvdkvaftgstdvgkiilela 290
QY 246 LKSNMKRVTTLELGEK-PTIVFNEAPLERQSGESAKDFSKF---GQIWVPPSCLLVQ--W 298
Db 291 sksnikavtleeeshhsfcedadvd-qavela-hfalffnqgqccagstrtfvhervy 348
QY 299 GNLAERFHGVRHGSFGGCGQWLGNQNPLEPKRT---HGPFVDKSQYDRVLGNIDVGKDT-A 354
Db 349 defvekaka-----ralkrnvgdppfksgigcggpqvdsdqfnkilkikhgveaga 398
QY 355 OLLTGVRKGDKGFAIEPTIVNPKPGSKIWFEEIFGVPVLSIKTFKTEEEAIEIANDTTY 414
Db 399 tlqaggdrlsgkgyyiqptvfsdvkddmliatdei fgpvqtilkfklddeviarannsr 458
QY 415 GLASVIYTKSLNRLVSSALETGGVSINFPFIPETQTTPFGMKQSGSGRELGEGLKAY 474
Db 459 glaagvftqnltdahrlmralrvgtvwinfcfdvidaslpfggkmsgigrekgyislnny 518
QY 475 LEPKTI 480
Db 519 lqvkvav 524

RESULT 14
ABB58519
ID ABB58519 standard; Protein; 909 AA.
XX
AC ABB58519;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 2349.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
XX
OS Drosophila melanogaster.
XX

PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX (PEKE) PE CORP NY.
PA
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
XX WPI; 2001-656860/75.
DR N-PSDB; ABL02622.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 2349; 2lpp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 909 AA;

Query Match 27.6%; Score 704; DB 22; Length 909;
Best Local Similarity 34.7%; Pred. No. 1.1e-58;
Matches 172; Conservative 89; Mismatches 196; Indels 38; Gaps 13;

QY 10 SELFINNEFVSSKGSERLTLTNPWDESTVATDVHVANAADVDSAVAASVQAVKKGPKWKF 69
Db 429 tqllingefvdaeaqrtilvnpntnee-vlckvacasatdvdkavraahsaf-ygswrql 486
QY 70 TGAQRAACMLKFADLAEKNAEKLARLESPTGRPPVSMITHFDIPNMVSVFRYYAGWADKI 129
Db 487 tprgrgqlmlnladlmerakeelatiesvdsgevtylalkthvgmsieawryfagwccki 546
QY 130 AGKTFPEDNGKPN-----WRVEPMGVCAGIASWNATFLYGVGKFIAPALAACGSFIFKASE 184
Db 547 qgntipvnparpnvnltftrkepigvcglitpwnyplmlswkmaaciaagntcclikpaq 606
QY 185 KSPLGVLGLAPLFAEAGFPFGVQFLTG-ARVTGEALASHMDIAKISFTFRSVGGGRAVKQ 243
Db 607 tcpltalkfaeltvragfpvgvinvlpkgsgdagqavadhelvrklgftgstpigkhmk 666
QY 244 ATLKSNMKRVTTLELGEK-PTIVFNEAPLER--QSGESAKDFSKFGQIWVPPSCLLVQWGN 300
Db 667 scadsnlkkcslelggksplilfadcdmdkavkhgmssvffnk-genciaagrlfve--- 722
QY 301 LAERKFH-----GVRHGSFGGCGQWLGNQNPLEPKRTHGPFVDKSQYDRVLGNIDVGK 351
Db 723 --drihdefirrvlkdirtmtig-----dpldrstahgpqnkhahfckllefcrrgv 772
QY 352 DT-AQLLTGVGRKGD-KGFAIEPTIFVNPKPGSKIWFEEIFGPVLSIKTFKTE--EEAIE 407
Db 773 degaklvvggcrvpnlkgyffttvftvntvtdmfdiaeesfgpimiliskfngsddslmq 832
QY 408 IANDTTYGLASVIYTKSLNRLVSSALETGGVSINFPFIPETQTTPFGMKQSGSGREL 467
Db 833 ranrteyglasgvftkdkigkalnfadrieaagtvfnvynktdvaapfggfkqsgygkdlg 892

QY 468 EEGLKAYLEPKTINI 482
:| | ||: | :
Db 893 gealneylktkcvtv 907

RESULT 15
AAG36240
ID AAG36240 standard; Protein: 424 AA.
XX
AC AAG36240;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 44384.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.

PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.

```
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match      27.5%; Score 700; DB 21; Length 424;
Best Local Similarity 39.3%; Pred. No. 8.3e-59;
Matches 169; Conservative 76; Mismatches 149; Indels 36; Gaps 9;

QY 70 TGAQRACMLKFADLAEKNAEKIARLESPTGRPVSMITHFDIPNMVSVFRYYAGWADKI 129
Db 2 tgferaklinkfadlieenieelakldavdggklfglgkyadipataghfrynagaadki 61

QY 130 AGKTFPEDNGKPNWRY---EPMGVCAGIASWNAFLYVGWKIAPALAAAGCSFIFKASEKS 186
Db 62 hgetl-kmtrqslfigytlkpeigvgvgniipwnfpsimfatkvapamaagctmvvkpaegt 120
```

```
QY 187 PLGVLGLAPLFAEAGFPFGVVQFLTG-ARVTGEALASHMDIAKISFTRSVGGGRAVKQAT 245
Db 121 slsalfyahlskeagipdgvlnivtfgfgstagaaiashmdvkvstfgstdvgrkimqaa 180
QY 246 LKSNMKRVTLELGEK-PTIVFNEAPLERQSGESAKD-----FSKFGQIWWPPSCLLVQW 298
Db 181 aasnkkvslelggksplliindadidk-----aadlallgcfnkgeicvassrvfvqe 235
QY 299 G-----NLAEKFHGVRHGSFGGCGRWLGQNPLEPKRTHGPFVVDKSOYDRVLGNIDVGK 351
Db 236 giydkvveklvek-----akdwtvgdpfdstarqqppqvdkrqfekilsyiehgk 284
QY 352 -DTAQLLTGVGRKKGDKGFAIEPTIFVNPKPGRSGKIWFEEIFGPVLSIKTFKTEEEAIEIAN 410
Db 285 negatiltggkaigdkgyfiqptifadvtdemkiyqdeifgpvmslmkfkveegikcan 344
QY 411 DTTYGLASVIYTKSLNRGLRVSSALETGVSINPFPIPETOTPFGGMKQSGSGRELGEEG 470
Db 345 ntkyglaagilsqdidlintvrsrikagiliwvncyfgfdldcpyggykmsgncresgmda 404
QY 471 LKAYLEPKTI 480
Db 405 ldnylgtksv 414
```

Search completed: April 27, 2002, 07:45:33
Job time: 8114 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic ~ nucleic search, using sw model

Run on: April 26, 2002, 20:50:00 ; Search time 5001.19 Seconds
(without alignments)
3950.968 Million cell updates/sec

Title: US-09-882-694A-4
Perfect score: 1464
Sequence: 1 atggttcttttcgctgacga.....tccacgtcaacatagagtga 1464

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- EST: *
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	232	15.8	2190	11	AK012213	Mus muscu
2	158.8	10.8	653	10	BG810006	mgct002xf
3	145	9.9	844	10	BF253896	HVSMef000
4	143.4	9.8	673	10	BI850371	imageqc_1
5	143.4	9.8	1024	9	AL135209	DKF2p762D
6	140.6	9.6	603	10	BE405431	WHE1216_C
7	139	9.5	1069	12	CNS0776W	T3 end of
8	138.2	9.4	660	10	BI955705	HVSMem002
9	137.4	9.4	1100	10	BM461228	AGENCOURT
10	131.8	9.0	1099	12	CNS0605B	T7 end of
11	131.2	9.0	656	9	BB626306	BB626306
12	130	8.9	395	10	BE428385	MTD006.D1
13	128.2	8.8	587	10	BE360262	DG1_62_B0
14	128	8.7	727	10	BF260792	BF260792
15	127.6	8.7	612	9	AW565878	HVSMef002
16	126	8.6	614	9	AV917894	LG1_352_H
17	124.6	8.5	674	9	AW243051	AV917894
						AW243051 xn28f12.x

18	124.2	8.5	508	9	AI158035	AI158035 ud26b09.r
19	124	8.5	636	9	AV929175	AV929175 AV929175
20	120.8	8.3	937	9	AL576537	AL576537 AL576537
21	120.8	8.3	993	9	AL564722	AL564722 AL564722
22	120.8	8.3	1047	12	CNS06NWY	AL407336 T7 end of
23	120	8.2	656	10	BG356773	BG356773 OV2_9_G07
24	120	8.2	1056	9	AL545580	AL545580 AL545580
25	119.8	8.2	736	9	AV721915	AV721915 AV721915
26	119.8	8.2	915	10	BI517895	BI517895 603041971
27	119.2	8.1	1014	9	AL532910	AL532910 AL532910
28	119	8.1	794	9	AL538753	AL538753 AL538753
29	118.6	8.1	795	10	BF261094	BF261094 HVSMef002
30	118.2	8.1	794	10	BG740760	BG740760 602632732
31	118.2	8.1	919	9	AI068478	AI068478 mgae0002c
32	118	8.1	643	10	BE362147	BE362147 DG1_84_D0
33	117	8.0	703	10	BE362145	BE362145 DG1_84_D0
34	117	8.0	723	9	AV721790	AV721790 AV721790
35	116.6	8.0	755	10	BG967081	BG967081 602833926
36	116.2	7.9	496	9	AA821260	AA821260 vs15a12.r
37	116	7.9	488	10	BG357200	BG357200 OV2_11_A0
38	116	7.9	740	10	BG644873	BG644873 EST506492
39	115.6	7.9	507	6	AW565284	AW565284 LG1_332_G
40	115.6	7.9	507	10	BM317596	BM317596 LGL_214_G
41	115.6	7.9	597	9	AI114304	AI114304 GH1111.3
42	115.6	7.9	630	9	AI455414	AI455414 LD23985.3
43	115.6	7.9	972	9	AI663535	AI663535 ui98a08.y
44	115	7.9	664	10	BG858411	BG858411 1024057E0
45	115	7.9	739	10	BG915019	BG915019 602813895

ALIGNMENTS

RESULT 1	AK012213	2190 bp	mRNA	linear	HTC 19-JAN-2002
LOCUS	Mus musculus	11 days embryo whole body cDNA, RIKEN full-length			
DEFINITION	enriched library, clone:2700007F14:homolog to ALDEHYDE DEHYDROGENASE, MITOCHONDRIAL X PRECURSOR (EC 1.2.1.3) (ALDH CLASS 2), full insert sequence.				
ACCESSION	AK012213				
VERSION	AK012213.1	GI:12848818			
KEYWORDS	HTC; CAP trapper.				
SOURCE	Mus musculus (strain:C57BL/6J) 11 days embryo cDNA to mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library clone:2700007F14.				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.				
AUTHORS	1 (sites)				
TITLE	Carninci, P. and Hayashizaki, Y.				
JOURNAL	High-efficiency full-length cDNA cloning				
MEDLINE	Meth. Enzymol. 303, 19-44 (1999)				
PUBMED	99279253				
REFERENCE	2 (sites)				
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.				
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes				
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)				
MEDLINE	20499374				
PUBMED	11042159				
REFERENCE	3 (sites)				
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.				
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer				


```
ALD4 ; aldehyde dehydrogenase, mitochondrial ]"
/evidence-not_experimental
<15. .>1034
/note="similar to AAA83769 [ Aldehyde dehydrogenase ] [
Pichia angusta ]"
/evidence-not_experimental
BASE COUNT      249 a   305 c   296 g   215 t       4 others
ORIGIN

Query Match          9.5%; Score 139; DB 12; Length 1069;
Best Local Similarity 51.4%; Pred. NO. 3.9e-26;
Matches 529; Conservative 1; Mismatches 469; Indels 30; Gaps 8;

QY 242 tcgcgacacctgcgcgagaagaacgcgcgagaagctcgctcgtctggagtcgctgccccaccg 301
   ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 16 TGCCCGAGCTGCTCGAGGCTCACAAGGATACCCCTCGCCGCAATTGAGTCTCTCGACAACG 75

QY 302 gtagaccggtgtcgatgatcactcatttcgacattccaaacatggtctccgtgtttcgc 361
   ||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 76 GTAAGSCTCTGCAGCTG--GCCAGGGTGACGTGGCGCTCGTGATCAACTACATCAGAT 132

QY 362 actatgcaggctggccgacagaagatcgccggaaagacattccccgaggacaaacggcaagc 421
   | ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 133 CGTGGCTGGCTGGCCGACAAAGCTGACCGGCCAGTACATCGACACGGGCGATACCCACC 192

QY 422 cgaattggcggtta---cgagccgatgggggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 478
   ||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 193 ACACCTACACCAAGAGAGAGAGCCCTATTGGAGTGTGTGGCCGACATCATCCCATGGAATTCC 252

QY 479 cttttctttacgtcgctgggaagatagccccccgcctcgccgcggctgctccttcattct 538
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 253 CTCCTTTGATGTGTCCTCGAAGGTGGTTCCTGCTCTGGCTGTCAGGTAACACTGTGTGTC 312

QY 539 tcaagcctcggagaaaatcccgcctggcgcttctgggcccctcgctcctctcttcgcagaag 598
   ||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 313 TGAAGACGGCCGAGAGACACGCCATTGTCTGCCCTGTACGGTTTCGACGTCGTCAAGGAGG 372

QY 599 ccggtatccctcctggagtcgtgcagttcctcactgcagagacagtg---acgggtgaag 655
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 373 CCGGAATCCCAGCCCGAGTGGTGAACATCGTGTCCGGATTCCGTAAGATCACGGGAGATG 432

QY 656 cattggcgtgcacatggacattgcgaagatcagcttcacaagatctgtcggcggtggcc 715
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 433 CCATTGCCAGCCACCCCTAAGATCAAGAAGCTTGCCCTTACCGGCTCCACCGCCACTGGAA 492

QY 716 gcgcgctcaagcaagcaactcaagtccaacatgaagcgcgtcactctagaaactggggg 775
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 493 GACACATCATGAAGGTTGCTGCCGACTCCAAACCTCAAGAAGTGACCCCTGGAGCTTGGAG 552

QY 776 aaaaagcaaccatcgtcttccaagaaagctcctctcg---aacggcagtcgggggaatcg 831
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 553 GAAAGTCTGCCAACATTTGTTTGACGACGCTGACGTCGAAGACCGCCATCAGCGCCCTCG 612

QY 832 --gcaaggatttctcaaaaattcggggcaaaatttgggtccccccctcctgtttgctagtgc 889
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 613 TTGCCGCTATCTTCTACAACTCCGAGAGAGGTGTGCTCTGCCGGATCCAGAGTGTTTGTCC 672

QY 890 aatggggaatttagcggagaaattccatggagtcctgcgtcattggtcatttggaggctgtc 949
   | ||||| ||| ||||| ||| ||| ||||| ||| ||| ||||| ||| ||||| ||| |||
Db 673 AGGAGGTTATTACGAGGAGATCTTCAGGGAA---TTCAAAGCTGCTGCAGAGACGCTC 728

QY 950 agataggcttggccagaacccattggaaccccaagaggagcagcatggtcccttcgtcgaca 1009
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 729 AAGGTGGG-----TGACCCCATCGATCTGTCCACTTTTCCAAGGTGCCCAAACTTCCC 780

QY 1010 agtcccagtagcacagagtcttgggtaaacattgacgttg---gcaaggataccgcgcagc 1066
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 781 AGATGCAATTGAACAAGATTCTCGGCTACGTCGACATCGCCACCACGGAAGGAGGTAGAC 840

QY 1067 tcctactggcgttggtagaaaggcgacaaaggattcgcgattgaaccgacgatatttg 1126
   || ||||| ||| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 841 TCATCACCGGAGDGNANGACTGGGGCAACAAGGGATTCTTCGTCGAAGCCCAACGATCTTTG 900
```

```
QY 1127 tcaatcccaaacaggcgagcaaaatttggtttgaggagatctttggcccccgcttgtcca 1186
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 901 CCGACGTGACCCCGACATGAGAGCATACAAGGAGGAGATCTTTGGACCATTTGCCGTGA 960

QY 1187 ttaagacgttcaagacggaagaagaggccattgagattgccaatgacacgacttatgggc 1246
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 961 TCACGCCATTCAAGACCCGCCGACGAGGTGATCGAGTTGGCTAACGACTCCGAATACGGAC 1020

QY 1247 tagcctcgg 1255
   | | | | |
Db 1021 TTGCTGCAG 1029

RESULT 8
BI955705
LOCUS
DEFINITION
    BI955705 660 bp mRNA linear EST 19-OCT-2001
    HVSMEm0024C12f Hordeum vulgare green seedling EST library
    HVCDNA0014 (Blumeria infected) Hordeum vulgare cDNA clone
    HVSMEm0024C12f, mRNA sequence.
ACCESSION
    BI955705
VERSION
    BI955705.1 GI:16302252
KEYWORDS
    EST.
SOURCE
    barley.
ORGANISM
    Hordeum vulgare
    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
    Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
    ; Triticeae; Hordeum.
REFERENCE
    1 (bases 1 to 660)
AUTHORS
    Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Chin,A., Begum,D.,
    Frisch,D., Atkins,M., Yu,Y., Henry,D., Palmer,M., Rambo,T., Simmons
    ,J., Oates,R. and Main,D.
TITLE
    Development of a genetically and physically anchored EST resource
    for barley genomics: Blumeria infected Morex (compatible) seedling
    cDNA library
JOURNAL
    Unpublished (2001)
COMMENT
    Contact: Wing RA
    Clemson University Genomics Institute
    Clemson University
    100 Jordan Hall, Clemson, SC 29634, USA
    Tel: 864 656 7288
    Fax: 864 656 4293
    Email: rwing@clemson.edu
    Total hg bases = 420
    Seq primer: AATTAAACCCCTCACTAAAGGG
    High quality sequence start: 10
    High quality sequence stop: 642.
FEATURES
    Location/Qualifiers
        1..660
            /organism="Hordeum vulgare"
            /cultivar="Morex"
            /db_xref="taxon:4513"
            /clone="HVSMEm0024C12f"
            /clone_lib="Hordeum vulgare green seedling EST library
            HVCDNA0014 (Blumeria infected)"
            /tissue_type="green seedling leaf"
            /lab_host="TJC121"
            /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
            XhoI; Morex (mia) plants were greenhouse grown in the R
            Wise lab at Iowa State University, Ames, IA; 7 day old
            green seedlings were infected with isolate 5874 of
            Blumeria graminis f. sp. hordei, and leaves were harvested
            24, 48 and 72 hr post-inoculation and snap frozen (Wise).
            In the TJ Close lab at the University of California,
            Riverside, total RNA was prepared from each sample pool,
            equal quantities of all three RNA pools were combined,
            poly(A) RNA was purified from the mixture, one primary
            unamplified cDNA library was made, and 1 million pfu were
            in vivo excised to give pBluescript SK(-) cDNA phagemids
            (Chin). phagemids were plated and picked at the Clemson
            University Genomics Institute (CUGI) (Begum, Palmer,
            Frisch, Atkins and Wing). Plasmid DNA preparations, DNA
            sequencing and sequence analysis were performed at CUGI
```


Db 512 -----ATGAGTTTGTGAAAGGAGTGTGAGCGAGCTAAGAAATATGTTCTTGGG 561

Qy 967 aaccattggaaccaagagacgatggtcccttcgcagacaagtccacgtacacaga 1026

Db 562 AATCCTCTGACCCAGGAATAATCAAGGCCCTCAGATTGACAAGGAACAACATGATAA 621

Qy 1027 gtcttggttaacattgacgtttggaag--gataccggcagctcctcactggttggt 1083

Db 622 ATACTCGATCTCATTTGAGAGTGGGAAGAAAGAGGAGCCAAACTGGAGTGTGTTGGAGGA 681

Qy 1084 agaaagggcgacaagggattcgcaattgaaccgacgatatattgtcaatcccaaacaggc 1143

Db 682 CGCTGGGGGACAAAGGCTTCTTTGTGCAGCCACACAGTGTCTCCAACGTTGACTGATGAG 741

Qy 1144 agcaaaatttggtttgaggagatcttttgccccctgtcttgctccattaaagacgttcaagacg 1203

Db 742 ATCGGCATTGCCAAAGAGGAGATATTGGACCAGTGCAACAATCATGAAGTTAAGTCT 801

Qy 1204 gaagaagaggccattgagattgccaatgacacgacttatgggctagcctcggtcatttat 1263

Db 802 GTAGATGATGTGATCAAGAGAGCAACAATACTACCTATGTTTACGACGAGGACTCTTC 861

Qy 1264 accaaatctctcaacaggggtctccgtgtctcgtcgccgctcgagaccggtg-gcgtctc 1322

Db 862 ACTAAGAC-CTGGATAAGGCCATCACTGTGTCATCTGCTCTGCAGGCTGGGTGNTGTG 920

Qy 1323 gatcaacttcccccttatccccgagacacaaaactccgtttggcgcatgaacaatcggg 1382

Db 921 GGTAACTGCTATATGATGNTGTCAGCCAGTGCCTTCCGCTCGGTGGGATTCAGATGTCTGG 980

Qy 1383 ctccaggcagagagctagcggaagagggctcaaggcgtacttggagcccaaga 1435

Db 981 GAATGGAAGAAACTGGGTGAACATGGCCTTTATGAAACCCTGAGCCTCAAAA 1033

RESULT 10

CNS06Q5B/c

LOCUS

DEFINITION T7 end of clone AW0AA002H11 of library AW0AA from strain CLIB 89 of Yarrowia lipolytica, genomic survey sequence.

ACCESSION AL410229

VERSION AL410229.1 GI:12178348

KEYWORDS GSS.

SOURCE Yarrowia lipolytica.

ORGANISM Yarrowia lipolytica

REFERENCE 1 (bases 1 to 1099)

AUTHORS Casaregola,S., Neuveglise,C., Lepingle,A., Bon,E., Feynerol,C., Artiguenave,F., Wincker,P. and Gaillardin,C.

TITLE Genomic exploration of the hemiascomycetous yeasts: 17. Yarrowia lipolytica

JOURNAL FEBS Lett. 487 (1), 95-100 (2000)

MEDLINE 20584727

REFERENCE 2 (bases 1 to 1099)

AUTHORS Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G., Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S., de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B., Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S., Saurin,W., Tekala,F., Tofano-Nioche,C., Wesolowski-Louvel,M., Wincker,P. and Weissenbach,J.

TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies

JOURNAL FEBS Lett. 487 (1), 3-12 (2000)

MEDLINE 20584711

REFERENCE 3 (bases 1 to 1099)

AUTHORS Genoscope.

TITLE Direct Submission

JOURNAL Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

COMMENT This GSS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces

exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

FEATURES

source

1. 1099

/organism="Yarrowia lipolytica"

/strain="CLIB 89"

/db_xref="taxon:4952"

/clone="AW0AA002H11"

/clone_lib="AW0AA"

/note="end : T7"

misc_feature

/note="similar to Saccharomyces cerevisiae ORF YMR169c [ALD3 ; strong similarity to aldehyde dehydrogenase] similar to Saccharomyces cerevisiae ORF YMR170c [ALD2 ; aldehyde dehydrogenase 2 (NAD+)]"

/evidence=not_experimental

misc_feature

/complement(<1. >822)

/note="similar to Saccharomyces cerevisiae ORF YER073w [ALD5 ; aldehyde dehydrogenase (NAD+), mitochondrial]"

/evidence=not_experimental

misc_feature

/complement(<4. >822)

/note="similar to Saccharomyces cerevisiae ORF YOR374w [ALD4 ; aldehyde dehydrogenase, mitochondrial]"

/evidence=not_experimental

BASE COUNT 244 a 288 c 318 g 246 t 3 others

ORIGIN

Query Match 9.0%; Score 131.8; DB 12; Length 1099;

Best Local Similarity 54.2%; Pred. No. 3.3e-24;

Matches 359; Conservative 0; Mismatches 288; Indels 15; Gaps 4;

Qy 133 gcaaacgcgccgatgtcgacagtgcagtagccgttcggtgcagcggtcaaaaagggc 192

Db 681 GCTACTGCTGAGGACGTGGATCGGGCAGTAGTTGCTGCGAAAGCGGTTCCAGA----- 625

Qy 193 ccatggaagaagtccaaggtgcacaaacgcgcggcggtgcgtgcattgaagttcgcggacctc 252

Db 626 -GCTGGCGAGATGTCGCCGGTGAGGAGCGTGCCAAAGTTGCTGGACAACCTTCATCAATCTG 568

Qy 253 gccgagaagaacgccgagaagctgcgtctgtgtgagtcgctgccccaccggtagaccggtg 312

Db 567 GTGTCCAAGAACCTCGACACGGTGGCTGCCATCGAGGCTCTCGATTNCGGCAAGCCCTTT 508

Qy 313 tcatgatcactcattcttcgacattccaaacatggtctccgtgttttcgctactatgcaggc 372

Db 507 CAGCTCAATGCTCGGGGTGACATCGCCGGCGGCGCTGCGCGTCTACAAGTACTACGCAGGG 448

Qy 373 tgggcccgaagaatcgccgggaaagacctttcccgaggacaaacggcaagc--cgaaattgg 429

Db 447 TGGCGGGACAAGGTGTTGGTAAGACCATTTCTCAACACCACCACCAAGAAAGCTGGCGTACACT 388

Qy 430 cgttacgagccgatgggggtgtgtgctggtgattgccagctggaacgcgacttttctttac 489

Db 387 CTTACAGAGCCCCATGGAGTGTGTGGTCAGATCATTCCTCGGAACATATCCGTTTCTGATG 328

Qy 490 gtccggtggaagatagcccccgccctcgccgcgggctgctccttcatcttcaaaagcctcg 549

Db 327 GCCCGTGGAGATTGCCCTGCAATTGCGGCTGGCAACCGTGGTGGTATGAAGCTCGCG 268

Qy 550 gagaaatccccgctggcggttctgggacctgcctctctcttcgcagaagccgattccct 609

Db 267 GAAACACACCCCTCTGTGGATGCTGTATCTGTGCAATCTGTTCAAGAGGCGCGGTTCCCT 208

Qy 610 cctggagtcgtgcagttcctcactggagcagcagtgacggttgagc---attggcgctcg 666

Db 207 CCGGAGTGATCAACATCTTCACTGGCCACGCGGCCAAGGCTGGCTGGCTGGCTGAG 148

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2002, 01:35:52 ; Search time 96.68 Seconds
(without alignments)
2781.733 Million cell updates/sec

Title: US-09-882-694A-4
Perfect score: 1464
Sequence: 1 atggttcttcgcctgacga.....tccacgtcaacatagagtga 1464

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 139433 seqs, 91850399 residues

Total number of hits satisfying chosen parameters: 278866

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_New:*
1: /cgn2_6/ptodata/1/pna/pct_NEW_COMB.seq:*
2: /cgn2_6/ptodata/1/pna/US06_NEW_COMB.seq:*
3: /cgn2_6/ptodata/1/pna/US07_NEW_COMB.seq:*
4: /cgn2_6/ptodata/1/pna/US08_NEW_COMB.seq:*
5: /cgn2_6/ptodata/1/pna/US09_NEW_COMB.seq:*
6: /cgn2_6/ptodata/1/pna/US10_NEW_COMB.seq:*
7: /cgn2_6/ptodata/1/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	ID	Description
1	191.6	13.1	2032	6	US-10-116-802-371
2	90	6.1	994	6	US-10-106-698-2021
3	42.8	2.9	245	5	US-09-975-254-19603
4	41	2.8	247	5	US-09-975-254-11784
5	39	2.7	285	5	US-09-975-254-24522
6	37.6	2.6	254	5	US-09-975-254-22618
7	33.6	2.3	777	1	PCT-US02-09944-15
8	33.4	2.3	18902	6	US-10-105-299-6372
9	32.6	2.2	243	5	US-09-975-254-9185
10	32.6	2.2	1327	6	US-10-102-806-29
11	32.2	2.2	1224	6	US-10-107-431-118
12	32.2	2.2	45055	6	US-10-107-431-277
13	32	2.2	243	5	US-09-975-254-8937
14	32	2.2	245	5	US-09-975-254-18708
15	32	2.2	6655	1	PCT-US02-09921-139
16	31.6	2.2	4912	1	PCT-US02-09809-28
17	31.2	2.1	203	5	US-09-975-254-7828
18	31.2	2.1	930	6	US-10-107-431-258
19	31.2	2.1	48221	6	US-10-107-431-280
20	31	2.1	864	6	US-10-103-313-268
21	31	2.1	894	5	US-09-623-791A-4
22	31	2.1	944	6	US-10-103-313-151
23	31	2.1	2105	5	US-09-623-791A-144
24	31	2.1	2332	6	US-10-116-802-214
25	30.6	2.1	210	5	US-09-540-210B-22756
26	30.6	2.1	1736	6	US-10-116-802-13

C	27	30.4	2.1	789	6	US-10-105-299-14976	Sequence 14976, A
	28	30.4	2.1	990	6	US-10-107-431-6	Sequence 6, Appli
	29	30.4	2.1	15240	6	US-10-107-431-278	Sequence 278, App
	30	30.4	2.1	22008	1	PCT-US02-09239-242	Sequence 242, App
	31	30.4	2.1	22008	6	US-10-105-299-6302	Sequence 6302, Ap
	32	30.4	2.1	78056	6	US-10-109-551-1	Sequence 1, Appli
	33	30.2	2.1	157	5	US-09-540-210B-5255	Sequence 5255, Ap
	34	30.2	2.1	249	5	US-09-975-254-14133	Sequence 14133, A
	35	30.2	2.1	254	5	US-09-975-254-29804	Sequence 29804, A
	36	30.2	2.1	1867	7	US-60-365-264-26	Sequence 26, Appl
	37	30	2.0	1518	6	US-10-105-299-1296	Sequence 1296, Ap
	38	30	2.0	2913	6	US-10-105-299-9405	Sequence 9405, Ap
	39	30	2.0	17967	6	US-10-103-313-650	Sequence 650, App
	40	29.8	2.0	256	5	US-09-540-210B-5883	Sequence 5883, Ap
	41	29.8	2.0	285	5	US-09-540-210B-13380	Sequence 13380, A
	42	29.8	2.0	302	5	US-09-540-210B-29113	Sequence 29113, A
	43	29.8	2.0	520	6	US-10-106-698-2579	Sequence 2579, Ap
	44	29.8	2.0	1392	6	US-10-105-299-2334	Sequence 2334, Ap
	45	29.8	2.0	3045	7	US-60-365-264-17	Sequence 17, Appl

ALIGNMENTS

RESULT 1
US-10-116-802-371
; Sequence 371, Application US/10116802
; GENERAL INFORMATION:
; APPLICANT: Amy Lasek
; TITLE OF INVENTION: GENES EXPRESSED IN LUNG CANCER
; FILE REFERENCE: PA-0045 US
; CURRENT APPLICATION NUMBER: US/10/116,802
; CURRENT FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: 60/281,593
; PRIOR FILING DATE: 2001-04-04
; NUMBER OF SEQ ID NOS: 519
; SOFTWARE: PERL Program
; SEQ ID NO 371
; LENGTH: 2032
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 3713867CB1
US-10-116-802-371

Query Match 13.1%; Score 191.6; DB 6; Length 2032;
Best Local Similarity 48.7%; Pred. No. 2.1e-47;
Matches 707; Conservative 0; Mismatches 724; Indels 21; Gaps 6;

Qy	12	gctgacgaatacaagagtgaactcttcatacaacaataatcgctctcctccaagggtc	71
Db	158	gccgaggtctctgcaaccagattttcataaacaataatgaatggcacgatgcgtcagcag	217
Qy	72	cgagagattaacgctcaagaacccgtgggacgaataccctgtgccactgatgttcacgt	131
Db	218	gaaacattccccaccgtcaatccgtccactggagaggtcatctgtcaggtagctgaagg	277
Qy	132	ggcaacgcggccgatgtcgacagtcagtagccgcttcggtgcaggcggtcaaaaagg	191
Db	278	ggacaa--ggaagatgtgacaaggcagtgaaagccgcccggcccttcagctggg	334
Qy	192	cccatggaagaagttcacaggtgcacaacgcgcggtgcatgcttaa---gttcgcgga	248
Db	335	ctcaccttggcgcgcgatggacgcacatcacacagggcggtgctgtaaccgctggccga	394
Qy	249	cctgcgcgagaagaacgcgcgagaagctcgctcgctggtgagtcgctgccaccggtagacc	308
Db	395	tcgatcgagcgggaccggaccctacctggcgcccttgagacccctggacaatggcaagcc	454
Qy	309	ggtgctgatgatcactcatttcgacattcccaaacatgggtctccggtgttcgctactatgc	368

Db 455 ctatgtcatctcctaccttggtggtgatttggacatggtcctcaaatgtctccggtattatgc 514

Qy 369 aggtctggccgacaaagatcgccggaaaagacctttccc--gaggacaaacggcaagccgaa 425

Db 515 cggctggctgataaagtaccacacgggaaaccatcccccattgcggagacttcttcagcta 574

Qy 426 ttggcgttacgagccgatgggggtgtgtgctggtgattgccagctggaaacgcgacttttct 485

Db 575 cacacgccatgaacctgtgggggtgtgcgggcagatcattccgtggaattcccgctcct 634

Qy 486 ttacgtcggctggaagatagccccccgcctcgccgctgctcctccttcattctcaaaagc 545

Db 635 gatcaagcatggaagctgggccagccttggcaactggaaacgtggtgtgatgaaggt 694

Qy 546 ctgggagaaatccccgcgtggggttcttgggcccctcgccgctcctctcttcgcagaaacggatt 605

Db 695 agctgagcagacacccctcacgcctctctatgtggccaacctgatcaaggaggtggctt 754

Qy 606 cctcctggagtcgtgcagttcctcactggagcacgagtgacgggtgaa--gcattggc 662

Db 755 tccccctggtgtgtcaacattgtgcctggatttggccccacggctggggccgccattgc 814

Qy 663 gtgcacatggacattgcgaagatcagcttcacaagatctgtcggcgtggccgcgcgcgt 722

Db 815 ctcccatgaggtgtggacaagtggcattcacaggctccactgagattggccgcgctaatt 874

Qy 723 caagcaagcaaacactcaagtccaacatgaagcgcgtcaactctagaacctgggggaaaaagcc 782

Db 875 ccaggttctgtctgggagcagcaacctcaagagagtgaccttggagctggggggggaagag 934

Qy 783 aaccatcgtcttcaacgaagctcctctctcgaaacggcagtcggggggaatcgggcaaggattt 842

Db 935 ccccaacatcatcatgtcagatgccgatatggttgggcccgtggaaacaggcccaacttgc 994

Qy 843 ctcaaaaattcgggcaaatttgggttccccccctcctctgttcttagtgcaatggggaaaaattt 902

Db 995 cctgttcttcaacca-----gggccagtgtctgtctgtcgcggctcccggaaccttcgtgca 1048

Qy 903 agcggagaaaattccatggagtcctcgtcatggtcatttggaggctgtcagagatggcttgg 962

Db 1049 ggaggacatctatgatgagtttgttgagcggagacttgcggggcccaagtctcgggttgg 1108

Qy 963 ccagaacccatttgaaacccaagaggacgcgcatggtcccttgcgtcgacaaagtcccaagtacga 1022

Db 1109 cgggaacccctttagatagcaagaccgagcaggggggccgcaggtggtgataaaactcagtttaa 1168

Qy 1023 cagagtccttgggtaaacattgacgttggcaag--gataccgcgcagctcctcacttgcgcgt 1079

Db 1169 gaagatcctcgtctacatacaacacggggaagcaaggggggcgaagctgctgtgtgtgtgg 1228

Qy 1080 tggtagaaaggcgacaaagggatttcgcgattgaaacgcgacgatatatttgtcaatcccaaac 1139

Db 1229 gggcatgtcgtgctgacgtggttacttccatccagcccaactgtgtttgagagtgtgcagga 1288

Qy 1140 aggcagcaaaaatttggtttgaggagatcttctggcccccgctcttgcatttaagacgttcaa 1199

Db 1289 tggcatgaccatcgccaaggaggagatcttcggggccagtgatgcagatcctgaagtcaa 1348

Qy 1200 gacggaagaagaggccattgagatttgccaatgacagacttatgggctagcctcggtcat 1259

Db 1349 gaccatagaggaggttgttgggagagaccaaaattccacgtacgggtgacggcgcagctgt 1408

Qy 1260 ttataccaaaatctctcaacaggggtctcctcgtgtctcgtcggcgtcgagaccggtggcgt 1319

Db 1409 cttaacaaaggatttggacaaggccaattacctgtccagggccctccagggcggaactgt 1468

Qy 1320 ctogatcaacttcccccttatcccccgagacacaaactccgttttggcggcatgaaacaact 1379

Db 1469 gtgggtcaactgctatgatgtgttggagccccagtcaccccttgggtgtacagatgtc 1528

Qy 1380 gggctcaggcagagagctaggcgaaagggtcgaaggcgtcacttggagccccagaccat 1439

Db 1529 ggggagtgggccgggagttgggcgagtagcgggctgcaggccatacactgaagtgaactgt 1588

Qy 1440 taatatccacgt 1451

Db 1589 cacagtcaaagt 1600

RESULT 2

US-10-106-698-2021/c

; Sequence 2021, Application US/10106698

; GENERAL INFORMATION:

; APPLICANT: Ruben et al.

; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypep

; FILE REFERENCE: PA005P1

; CURRENT APPLICATION NUMBER: US/10/106,698

; CURRENT FILING DATE: 2002-03-27

; PRIOR APPLICATION NUMBER: PCT/US00/26524

; PRIOR FILING DATE: 2000-09-28

; PRIOR APPLICATION NUMBER: US 60/157,137

; PRIOR FILING DATE: 1999-09-29

; PRIOR APPLICATION NUMBER: US 60/163,280

; PRIOR FILING DATE: 1999-11-03

; NUMBER OF SEQ ID NOS: 8564

; SOFTWARE: PatentIn Ver. 3.0

; SEQ ID NO 2021

; LENGTH: 994

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-106-698-2021

Query Match 6.1%; Score 90; DB 6; Length 994;

Best Local Similarity 55.9%; Pred. No. 2.4e-17;

Matches 171; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

Qy 1146 caaaaatttggttgaggagatctttggccccgctctgtccattaaagacgttcaagacgga 1205

Db 853 CAAGATCGCCAAGGAGGAGATCTTCGGGCCAGTGATCGAGATCCTGAAGTTCAAGACCAT 794

Qy 1206 agaagaggccattgagattgccaatgacacgacttatgggctagcctcggctcattatac 1265

Db 793 AGAGGAGGTTGTGGGAGAGCCCAACAATTCCACGTACGGGCTGCCGCAGCTGTCTTCAC 734

Qy 1266 caaatctctcaacaggggtctcctcgtgtcgtcgcgtcgcagaccgggtggtctcgtat 1325

Db 733 AAAGGATTGGACAAGGCCCAATTACCTGTCCCAGGCCCTCCAGGGGCACACTGTGTGGGT 674

Qy 1326 caacttccccctttatccccgagagacacaaactccgttttggcgcatgaaacaactcgggtc 1385

Db 673 CAACTGCTATGATGTGTTGGAGCCCCAGTCACCCCTTTGGTGGCTACAAGATGTCCGGGSAG 614

Qy 1386 aggcagagagctaggcgaagaagggtcgaaggcgtacttggagccccaaagaccattatat 1445

Db 613 TGGCCGGGAGTTGGCGGAGTACGGGCTGCAGGCATACACTGAAGTGAAAACCTGTCACAGT 554

Qy 1446 ccacgt 1451

Db 553 CAAAGT 548

RESULT 3

US-09-975-254-19603

; Sequence 19603, Application US/09975254

; GENERAL INFORMATION:

; APPLICANT: Byrum, Joseph R.

; APPLICANT: Heck, Gregory R.

; APPLICANT: La Rosa, Thomas J.

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

; FILE REFERENCE: 38-21(15309)B

; CURRENT APPLICATION NUMBER: US/09/975,254

; CURRENT FILING DATE: 2001-10-12

; PRIOR APPLICATION NUMBER: US/09/263,191

; PRIOR FILING DATE: 1999-03-05

Db 146 ttggaccggtttaccagtaattaggatcaattcagttgaagaaggaatccatcattgt 205
QY 1228 aatgacacgacttatgggct 1247
Db 206 aatgctagcaactttggact 225

RESULT 7
PCT-US02-09944-15
; Sequence 15, Application PC/TUS0209944
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: DAFFO, Abel
; APPLICANT: JONES, Anissa L.
; APPLICANT: TRAN, Alanna-Phung B.
; APPLICANT: DAHL, Christopher R.
; APPLICANT: GIETZEN, Darryl
; APPLICANT: CHINN, Joyce
; APPLICANT: DUFOUR, Gerard E.
; APPLICANT: HILLMAN, Jennifer L.
; APPLICANT: YU, Jimmy Y.
; APPLICANT: TUASON, Olivia
; APPLICANT: YAP, Pierre E.
; APPLICANT: AMSHEY, Stefan R.
; APPLICANT: DAUGHERTY, Sean C.
; APPLICANT: DAM, Tam C.
; APPLICANT: LIU, Tommy F.
; APPLICANT: NGUYEN, Duy-Viet An
; APPLICANT: KLEEFELD, Yael
; APPLICANT: GERSTIN, JR., Edward H.
; APPLICANT: PERALTA, Careyna H.
; APPLICANT: DAVID, Marie H.
; APPLICANT: LEWIS, Samantha A.
; APPLICANT: CHEN, Alice J.
; APPLICANT: PANZER, Scott R.
; APPLICANT: HARRIS, Bernard
; APPLICANT: FLORES, Vincent
; APPLICANT: MARWAHA, Rakesh
; APPLICANT: LO, Audrey
; APPLICANT: LAN, Ruth Y.
; APPLICANT: URASHKA, Michael E.
; TITLE OF INVENTION: MOLECULES FOR DISEASE DETECTION AND TREATMENT
; FILE REFERENCE: PT-1231 PCT
; CURRENT APPLICATION NUMBER: PCT/US02/09944
; PRIOR FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: 60/280,067; 60/279,619; 60/280,068; 60/291,280; 60/291,849;
; 60/291,829; 60/299,428; 60/300,001; 60/299,776
; PRIOR FILING DATE: 2001-03-29; 2001-03-28; 2001-03-29; 2001-05-16; 2001-05-17;
; 2001-05-17; 2001-06-19; 2001-06-20; 2001-06-20
; NUMBER OF SEQ ID NOS: 792
; SOFTWARE: PERL Program
; SEQ ID NO 15
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: LG:220648.6:2001MAR30
PCT-US02-09944-15

Query Match 2.3%; Score 33.6; DB 1; Length 777;
Best Local Similarity 53.9%; Pred. No. 1.3;
Matches 69; Conservative 0; Mismatches 59; Indels 0; Gaps 0;
QY 464 ccagctggacgcgactttcttaccgtcggtggagatagcccccgcctcgccg 523
Db 442 ccgcgtcttcgagactgtgtatgatggctactccaaagcacgcgtccaccacctctgct 501
QY 524 gctgctccttcattctcaaaagcctggagaaaatccccgcgtggcgcttctgggctcgctc 583
Db 502 gccagtagcactctgtgaacgcgcacccggccttgccaccgcccgccacctggccctggccc 561

QY 584 ctctcttc 591
Db 562 ccactcctc 569
RESULT 8
US-10-105-299-6372
; Sequence 6372, Application US/10105299
; GENERAL INFORMATION:
; APPLICANT: Rosen, et. al
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS950
; CURRENT APPLICATION NUMBER: US/10/105,299
; CURRENT FILING DATE: 2002-03-26
; NUMBER OF SEQ ID NOS: 15197
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6372
; LENGTH: 18902
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4483)..(4483)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-105-299-6372

Query Match 2.3%; Score 33.4; DB 6; Length 18902;
Best Local Similarity 47.0%; Pred. No. 4.7;
Matches 103; Conservative 0; Mismatches 116; Indels 0; Gaps 0;
QY 370 ggctgggcccgaagatcgccggaagaccttcccgaggacaaacggcgaattgg 429
Db 12165 ggctgggagacagaggtcttgggaatggcagttcttaggggtggggtggaatg 12224
QY 430 cgttacgagccgatgggggtgtgtgctggtattgccagctggaaacgcgactttcttac 489
Db 12225 ctctccaaagccagagggggccggcctggatttggaggtggacagggggtcttcaaac 12284
QY 490 gtcggctggaagatagcccccgcctcgccgcggctgctctcttcaaaagcctcg 549
Db 12285 cccctaccatactcattccagggtcctatgtgtccagctgtcagagcagggtcccaatcctgc 12344
QY 550 gagaaatccccgcgtggcggttcttggccctcgctcctctc 588
Db 12345 tccccattaccttggaacatcctgtacaggattccctc 12383

RESULT 9
US-09-975-254-9185
; Sequence 9185, Application US/09975254
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: Heck, Gregory R.
; APPLICANT: La Rosa, Thomas J.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(15309)B
; CURRENT APPLICATION NUMBER: US/09/975,254
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US/09/263,191
; PRIOR FILING DATE: 1999-03-05
; NUMBER OF SEQ ID NOS: 31255
; SEQ ID NO 9185
; LENGTH: 243
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: 700755553H1
US-09-975-254-9185

Matches	71;	Conservative	0;	Mismatches	65;	Indels	0;	Gaps	0;
Qy	507	ccccgccctcgccgcccgtgctccttcattcctcgaagccctcgagagaaatccccgctggg	566						
Db	6065	CCGCGGCCCCACCTATGCTGTGACCTTTAACTTGAAGACCAAGAGAGAGCCAGCGAAGTG	5006						
Qy	567	cgttcctgggctcgctcctccttccttcgcagaaagccggattccctcctcgagtcgtgcagtt	526						
Db	6005	GATCCTGGCAGCCGTGGCGCTGCTCCTACAGGTTTAGCTCGCTCCTGCCTCAGAGCCAC	5946						
Qy	627	cctcactggagcacga	642						
Db	5945	ACTCCCTGGGCGCTGGA	5930						

Search completed: April 27, 2002, 01:36:31
Job time: 17124 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2002, 01:33:16 ; Search time 8172.36 Seconds
(without alignments)
3875.685 Million cell updates/sec

Title: US-09-882-694A-4
Perfect score: 1464
Sequence: 1 atggttcttgcctgacga.....tccacgtcaacatagagtga 1464

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 21979536 seqs, 10817449327 residues 43959072
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_Main:*

1:	/cgn2_6/ptodata/2/pna/PCTUS_COMB.seq:*
2:	/cgn2_6/ptodata/2/pna/US06_COMB.seq:*
3:	/cgn2_6/ptodata/2/pna/US07_COMB.seq:*
4:	/cgn2_6/ptodata/2/pna/US080_COMB.seq:*
5:	/cgn2_6/ptodata/2/pna/US081_COMB.seq:*
6:	/cgn2_6/ptodata/2/pna/US082_COMB.seq:*
7:	/cgn2_6/ptodata/2/pna/US083_COMB.seq:*
8:	/cgn2_6/ptodata/2/pna/US084_COMB.seq:*
9:	/cgn2_6/ptodata/2/pna/US085_COMB.seq:*
10:	/cgn2_6/ptodata/2/pna/US086_COMB.seq:*
11:	/cgn2_6/ptodata/2/pna/US087_COMB.seq:*
12:	/cgn2_6/ptodata/2/pna/US088_COMB.seq:*
13:	/cgn2_6/ptodata/2/pna/US089_COMB.seq:*
14:	/cgn2_6/ptodata/2/pna/US090_COMB.seq:*
15:	/cgn2_6/ptodata/2/pna/US091_COMB.seq:*
16:	/cgn2_6/ptodata/2/pna/US092_COMB.seq:*
17:	/cgn2_6/ptodata/2/pna/US093_COMB.seq:*
18:	/cgn2_6/ptodata/2/pna/US094_COMB.seq:*
19:	/cgn2_6/ptodata/2/pna/US095A_COMB.seq:*
20:	/cgn2_6/ptodata/2/pna/US095B_COMB.seq:*
21:	/cgn2_6/ptodata/2/pna/US095C_COMB.seq:*
22:	/cgn2_6/ptodata/2/pna/US095D_COMB.seq:*
23:	/cgn2_6/ptodata/2/pna/US096A_COMB.seq:*
24:	/cgn2_6/ptodata/2/pna/US096B_COMB.seq:*
25:	/cgn2_6/ptodata/2/pna/US096C_COMB.seq:*
26:	/cgn2_6/ptodata/2/pna/US096D_COMB.seq:*
27:	/cgn2_6/ptodata/2/pna/US096E_COMB.seq:*
28:	/cgn2_6/ptodata/2/pna/US097A_COMB.seq:*
29:	/cgn2_6/ptodata/2/pna/US097B_COMB.seq:*
30:	/cgn2_6/ptodata/2/pna/US097C_COMB.seq:*
31:	/cgn2_6/ptodata/2/pna/US098A_COMB.seq:*
32:	/cgn2_6/ptodata/2/pna/US098B_COMB.seq:*
33:	/cgn2_6/ptodata/2/pna/US098C_COMB.seq:*
34:	/cgn2_6/ptodata/2/pna/US099A_COMB.seq:*
35:	/cgn2_6/ptodata/2/pna/US099B_COMB.seq:*
36:	/cgn2_6/ptodata/2/pna/US099C_COMB.seq:*
37:	/cgn2_6/ptodata/2/pna/US100_COMB.seq:*
38:	/cgn2_6/ptodata/2/pna/US101_COMB.seq:*
39:	/cgn2_6/ptodata/2/pna/US6000_COMB.seq:*
40:	/cgn2_6/ptodata/2/pna/US6001_COMB.seq:*
41:	/cgn2_6/ptodata/2/pna/US6002_COMB.seq:*
42:	/cgn2_6/ptodata/2/pna/US6003_COMB.seq:*
43:	/cgn2_6/ptodata/2/pna/US6004_COMB.seq:*

44:	/cgn2_6/ptodata/2/pna/US6005_COMB.seq:*
45:	/cgn2_6/ptodata/2/pna/US6006_COMB.seq:*
46:	/cgn2_6/ptodata/2/pna/US6007_COMB.seq:*
47:	/cgn2_6/ptodata/2/pna/US6008_COMB.seq:*
48:	/cgn2_6/ptodata/2/pna/US6009_COMB.seq:*
49:	/cgn2_6/ptodata/2/pna/US6010_COMB.seq:*
50:	/cgn2_6/ptodata/2/pna/US6011_COMB.seq:*
51:	/cgn2_6/ptodata/2/pna/US6012_COMB.seq:*
52:	/cgn2_6/ptodata/2/pna/US6013_COMB.seq:*
53:	/cgn2_6/ptodata/2/pna/US6014_COMB.seq:*
54:	/cgn2_6/ptodata/2/pna/US6015_COMB.seq:*
55:	/cgn2_6/ptodata/2/pna/US6016_COMB.seq:*
56:	/cgn2_6/ptodata/2/pna/US6017_COMB.seq:*
57:	/cgn2_6/ptodata/2/pna/US6018_COMB.seq:*
58:	/cgn2_6/ptodata/2/pna/US6019_COMB.seq:*
59:	/cgn2_6/ptodata/2/pna/US6020_COMB.seq:*
60:	/cgn2_6/ptodata/2/pna/US6021_COMB.seq:*
61:	/cgn2_6/ptodata/2/pna/US6022_COMB.seq:*
62:	/cgn2_6/ptodata/2/pna/US6023_COMB.seq:*
63:	/cgn2_6/ptodata/2/pna/US6024_COMB.seq:*
64:	/cgn2_6/ptodata/2/pna/US6025_COMB.seq:*
65:	/cgn2_6/ptodata/2/pna/US6026_COMB.seq:*
66:	/cgn2_6/ptodata/2/pna/US6027_COMB.seq:*
67:	/cgn2_6/ptodata/2/pna/US6028_COMB.seq:*
68:	/cgn2_6/ptodata/2/pna/US6029_COMB.seq:*
69:	/cgn2_6/ptodata/2/pna/US6030_COMB.seq:*
70:	/cgn2_6/ptodata/2/pna/US6031_COMB.seq:*
71:	/cgn2_6/ptodata/2/pna/US6032_COMB.seq:*
72:	/cgn2_6/ptodata/2/pna/US6033_COMB.seq:*
73:	/cgn2_6/ptodata/2/pna/US6034_COMB.seq:*
74:	/cgn2_6/ptodata/2/pna/US6035_COMB.seq:*
75:	/cgn2_6/ptodata/2/pna/US6036_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1464	100.0	1464	17	US-09-351-224-4	Sequence 4, Appli
2	1464	100.0	1464	17	US-09-351-224E-4	Sequence 4, Appli
3	1464	100.0	1464	17	US-09-351-823-4	Sequence 4, Appli
4	1464	100.0	1464	26	US-09-677-488-4	Sequence 4, Appli
5	1464	100.0	1464	26	US-09-677-488A-4	Sequence 4, Appli
6	1464	100.0	1464	26	US-09-677-682-4	Sequence 4, Appli
7	1464	100.0	1464	26	US-09-677-682A-4	Sequence 4, Appli
8	1464	100.0	1464	26	US-09-677-682B-4	Sequence 4, Appli
9	1464	100.0	1464	33	US-09-882-694-4	Sequence 4, Appli
10	1464	100.0	1464	33	US-09-882-694A-4	Sequence 1, Appli
11	267.6	18.3	1491	10	US-08-612-853-1	Sequence 27986, A
12	250.6	17.1	1491	18	US-09-404-520-27986	Sequence 5519, Ap
13	245.8	16.8	2022	20	US-09-533-559-5519	Sequence 14, Appl
14	245.4	16.8	1710	20	US-09-533-559-14	Sequence 89, Appl
15	241.6	16.5	9589	1	PCT-US01-17253-89	Sequence 419, App
16	240.4	16.4	1551	61	US-60-226-176-419	Sequence 419, App
17	240.4	16.4	1551	62	US-60-233-468-419	Sequence 419, App
18	240.4	16.4	1551	70	US-60-313-371-419	Sequence 418, App
19	240.4	16.4	2041	61	US-60-226-176-418	Sequence 418, App
20	240.4	16.4	2041	62	US-60-233-468-418	Sequence 418, App
21	240.4	16.4	2041	70	US-60-313-371-418	Sequence 6928, Ap
22	240.4	16.4	2288	60	US-60-213-360-6928	Sequence 11821, A
23	240.4	16.4	2308	56	US-60-172-373-11821	Sequence 7895, Ap
24	240.4	16.4	2552	25	US-09-644-867-7895	Sequence 9732, Ap
25	240.4	16.4	2552	25	US-09-644-871-9732	Sequence 9381, Ap
26	240.4	16.4	2552	25	US-09-652-126-9381	Sequence 10113, A
27	240.4	16.4	2552	25	US-09-652-128-10113	Sequence 10185, A
28	240.4	16.4	2552	25	US-09-652-911-10185	Sequence 9132, Ap
29	240.4	16.4	2552	25	US-09-652-914-9132	Sequence 9779, Ap
30	240.4	16.4	2552	25	US-09-652-916-9779	Sequence 12416, A
31	240.4	16.4	2552	27	US-09-698-010-12416	

; Sequence 4, Application US/09351224E
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; TITLE OF INVENTION: Detoxification
; FILE REFERENCE: 5718-111
; CURRENT APPLICATION NUMBER: US/09/351,224E
; CURRENT FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1464
; TYPE: DNA
; ORGANISM: Exophiala spinifera
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: aldehyde dehydrogenase, fully spliced cDNA
US-09-351-224E-4

Query Match 100.0%; Score 1464; DB 17; Length 1464;
Best Local Similarity 100.0%; Pred. NO. 0;
Matches 1464; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggttcttcgacctgacgaatacaagaagtgaactcttcatcaacaatgaattgcttcc 60
Db 1 atggttcttcgacctgacgaatacaagaagtgaactcttcatcaacaatgaattgcttcc 60

QY 61 tccaaggggtccgagagattaaacgtcagcaaacccgtggagcaaatccaccgttgcact 120
Db 61 tccaaggggtccgagagattaaacgtcagcaaacccgtggagcaaatccaccgttgcact 120

QY 121 gatgttcacgtggccaaacgcccagatgtcgacagtgacagtagccgcttcggtgagcg 180
Db 121 gatgttcacgtggccaaacgcccagatgtcgacagtgacagtagccgcttcggtgagcg 180

QY 181 gtcaaaaaagggcccatggaagaagttaacaggtgcacaggtgcacaaacgcccgtatga 240
Db 181 gtcaaaaaagggcccatggaagaagttaacaggtgcacaggtgcacaaacgcccgtatga 240

QY 241 ttcgaggacctgcgagagaagaacgcccagagagctgctgctgctgctgctgctgctgct 300
Db 241 ttcgaggacctgcgagagaagaacgcccagagagctgctgctgctgctgctgctgctgct 300

QY 301 ggtagaccggtgctgatgacactaatcttcacattcccaacatggtctcctgcttctcgc 360
Db 301 ggtagaccggtgctgatgacactaatcttcacattcccaacatggtctcctgcttctcgc 360

QY 361 tactatgcaggctgggcccagacagatcgccggaagacatttcccagaggacaaacggcaag 420
Db 361 tactatgcaggctgggcccagacagatcgccggaagacatttcccagaggacaaacggcaag 420

QY 421 ccgaattggcgttacgagccgatgggggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 480
Db 421 ccgaattggcgttacgagccgatgggggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 480

QY 481 ttctttacgtgggtggaagatagccccccctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 540
Db 481 ttctttacgtgggtggaagatagccccccctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 540

QY 541 aaagcctcggagaataccccgctggggttcttggcctcgtcgtcgtcgtcgtcgtcgtcgt 600
Db 541 aaagcctcggagaataccccgctggggttcttggcctcgtcgtcgtcgtcgtcgtcgtcgt 600

QY 601 ggattccctcctgagtcgtgcagttcctcactggagcacgagtgacgggtgaagcattg 660
Db 601 ggattccctcctgagtcgtgcagttcctcactggagcacgagtgacgggtgaagcattg 660

QY 661 gcgtcgacatggacattcggaagatcagcttcaagaatctgtcggcggtggccgccc 720
Db 661 gcgtcgacatggacattcggaagatcagcttcaagaatctgtcggcggtggccgccc 720

QY 721 gtcaagcaagcaaacactcaagtcacacatgaagcgcgtcactctagaaactggggaaaaa 780
Db 721 gtcaagcaagcaaacactcaagtcacacatgaagcgcgtcactctagaaactggggaaaaa 780

QY 781 ccaaccatcgtcttcaacgaagctcctctcgaacgagcagtcgggggaatcgcaaaaggat 840
Db 781 ccaaccatcgtcttcaacgaagctcctctcgaacgagcagtcgggggaatcgcaaaaggat 840

QY 841 ttctcaaaaattcgggcaaaaatttgggtccccctcctcgttctgtagtgcattggggaaat 900
Db 841 ttctcaaaaattcgggcaaaaatttgggtccccctcctcgttctgtagtgcattggggaaat 900

QY 901 tttagcgagaaaattccatggagtcctcgtcatggtcatttggaggtcgtcagagatggctt 960
Db 901 tttagcgagaaaattccatggagtcctcgtcatggtcatttggaggtcgtcagagatggctt 960

QY 961 ggcagaaacccattggaaacccaaagagagcagtcgtcctcctcgtcgaacagtcctcagtac 1020
Db 961 ggcagaaacccattggaaacccaaagagagcagtcgtcctcctcgtcgaacagtcctcagtac 1020

QY 1021 gacagatcttgggttaacattgaacgttggcaaggtacacgcgcagctcctcactggcgctt 1080
Db 1021 gacagatcttgggttaacattgaacgttggcaaggtacacgcgcagctcctcactggcgctt 1080

QY 1081 ggtagaagggcgacaaaggtattcgcgattgaacccagacatatttgcattcccaaacca 1140
Db 1081 ggtagaagggcgacaaaggtattcgcgattgaacccagacatatttgcattcccaaacca 1140

QY 1141 ggcagcaaaaatttgggttggagagatcttggcccccttgcatttgcatttgcatttgcatt 1200
Db 1141 ggcagcaaaaatttgggttggagagatcttggcccccttgcatttgcatttgcatttgcatt 1200

QY 1201 acggaagagagggccattgagattgccaatgacacgacttatgggctagcctcgtcatt 1260
Db 1201 acggaagagagggccattgagattgccaatgacacgacttatgggctagcctcgtcatt 1260

QY 1261 tataccaaatctctcaacaggggtctcctcgtgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 1320
Db 1261 tataccaaatctctcaacaggggtctcctcgtgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 1320

QY 1321 tcgatacacttcccttatacccgagacacaaactcgttggcgcatgaaacaatcg 1380
Db 1321 tcgatacacttcccttatacccgagacacaaactcgttggcgcatgaaacaatcg 1380

QY 1381 ggctcaggcagagagtaggcgaagaagggctcaaggcgtacttgagccccagaccatt 1440
Db 1381 ggctcaggcagagagtaggcgaagaagggctcaaggcgtacttgagccccagaccatt 1440

QY 1441 aatatccacgtcaacatagatga 1464
Db 1441 aatatccacgtcaacatagatga 1464

RESULT 3
US-09-351-823-4
; Sequence 4, Application US/09351823
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jon
; TITLE OF INVENTION: Compositions and Methods for Fumonisin Detoxification
; FILE REFERENCE: 5718-56
; CURRENT APPLICATION NUMBER: US/09/351,823
; CURRENT FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1464
; TYPE: DNA
; ORGANISM: Exophiala spinifera
; FEATURE:


```
QY 181 gtcaaaaaggccccatggaagaagtccacaggtgcacaacgcgcggcgctgcatgcttaag 240
Db 181 gtcaaaaaggccccatggaagaagtccacaggtgcacaacgcgcggcgctgcatgcttaag 240

QY 241 ttcgcgacctgcgcgagaagaacgcgcgagaagctcgtcgtctgagtcgctgccacc 300
Db 241 ttcgcgacctgcgcgagaagaacgcgcgagaagctcgtcgtctgagtcgctgccacc 300

QY 301 ggtagaccggtgtcgatgacactcatttcgacattccaaacatggtctccgtgtttcgc 360
Db 301 ggtagaccggtgtcgatgacactcatttcgacattccaaacatggtctccgtgtttcgc 360

QY 361 tactatgcaggctggccgacaaagatgcgcggaaagacctttcccgaggacaacggcaag 420
Db 361 tactatgcaggctggccgacaaagatgcgcggaaagacctttcccgaggacaacggcaag 420

QY 421 ccgaattggcgttacgagccgatgggggtgtgtgctggtattgccagctggaacgcgact 480
Db 421 ccgaattggcgttacgagccgatgggggtgtgtgctggtattgccagctggaacgcgact 480

QY 481 ttctttacgtcggctggaagatagcccccgccctgcgcggctgctccttcattcttc 540
Db 481 ttctttacgtcggctggaagatagcccccgccctgcgcggctgctccttcattcttc 540

QY 541 aaagcctcggagaaatccccgctggcgctttctggccctgcctcctctcttcgcagaagcc 600
Db 541 aaagcctcggagaaatccccgctggcgctttctggccctgcctcctctcttcgcagaagcc 600

QY 601 ggattccctcctggagtcgctgcagttcctcactcaggagcacgagtgacgggtgaagcattg 660
Db 601 ggattccctcctggagtcgctgcagttcctcactcaggagcacgagtgacgggtgaagcattg 660

QY 661 gcgtcgcacatggacattgcgaagatcagcttcacaagatctgtcggcggtgcccgcgcc 720
Db 661 gcgtcgcacatggacattgcgaagatcagcttcacaagatctgtcggcggtgcccgcgcc 720

QY 721 gtcaagcaagcaacactcaagtccaacatgaagcgctcactctagaaactgggggaaaaag 780
Db 721 gtcaagcaagcaacactcaagtccaacatgaagcgctcactctagaaactgggggaaaaag 780

QY 781 ccaaccatcgtcttcaacgaagctcctctcgaacggcagtcgggggaatcggcaaaaggat 840
Db 781 ccaaccatcgtcttcaacgaagctcctctcgaacggcagtcgggggaatcggcaaaaggat 840

QY 841 ttctcaaaaattcgggcaaaatttgggtcccccctcctctgttctgtagtgcaatggggaaaat 900
Db 841 ttctcaaaaattcgggcaaaatttgggtcccccctcctctgttctgtagtgcaatggggaaaat 900

QY 901 ttagcggagaaaaattccatggagtcctcgtcatggtcatttggaggctgtcagagatggctt 960
Db 901 ttagcggagaaaaattccatggagtcctcgtcatggtcatttggaggctgtcagagatggctt 960

QY 961 ggccagaacccattggaaacccaagaggacgcatggtcccttcgtcgcacaagtccacgtac 1020
Db 961 ggccagaacccattggaaacccaagaggacgcatggtcccttcgtcgcacaagtccacgtac 1020

QY 1021 gacagagtcttgggttaacattgacgttggcaaggatataccgcgcagctcctcactggcgtt 1080
Db 1021 gacagagtcttgggttaacattgacgttggcaaggatataccgcgcagctcctcactggcgtt 1080

QY 1081 ggtagaaaggcgacaaggattcgcgattgaaacgacgatatttgcattatcccaaacca 1140
Db 1081 ggtagaaaggcgacaaggattcgcgattgaaacgacgatatttgcattatcccaaacca 1140

QY 1141 ggcagcaaaaatttgggttgaggagatcttggcccgctcttgcattatgaacgcttcaag 1200
Db 1141 ggcagcaaaaatttgggttgaggagatcttggcccgctcttgcattatgaacgcttcaag 1200

QY 1201 acggaagaagaggccattgagattgccaatgacacgacttatgggctagcctcggtcatt 1260
Db 1201 acggaagaagaggccattgagattgccaatgacacgacttatgggctagcctcggtcatt 1260
```

```
QY 1261 tatacaaaatctctcaacagggggtctctccgtgtctcgttcggcgctcgagaccgggtggcgtc 1320
Db 1261 tatacaaaatctctcaacagggggtctctccgtgtctcgttcggcgctcgagaccgggtggcgtc 1320

QY 1321 tcgatcaacttccctttatcccccgagacacaaaactccgtttggcggcatgaaacaactcg 1380
Db 1321 tcgatcaacttccctttatcccccgagacacaaaactccgtttggcggcatgaaacaactcg 1380

QY 1381 ggctcaggcagagagctaggcgaagaagggtcgaaggctcacttggagcccagaccatt 1440
Db 1381 ggctcaggcagagagctaggcgaagaagggtcgaaggctcacttggagcccagaccatt 1440

QY 1441 aatatccacgtcaacatagagtga 1464
Db 1441 aatatccacgtcaacatagagtga 1464

RESULT 5
US-09-677-488A-4
; Sequence 4, Application US/09677488A
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; FILE REFERENCE: 35718/204100
; CURRENT APPLICATION NUMBER: US/09/677,488A
; CURRENT FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1464
; TYPE: DNA
; ORGANISM: Exophiala spinifera
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)..(0)
; OTHER INFORMATION: aldehyde dehydrogenase, fully spliced cDNA
US-09-677-488A-4

Query Match 100.0%; Score 1464; DB 26; Length 1464;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1464; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggttcttttcgcctgacgaatacaagagtgaactttcatcaacaatgaattcgtctcc 60
Db 1 atggttcttttcgcctgacgaatacaagagtgaactttcatcaacaatgaattcgtctcc 60

QY 61 tccaaagggtccgagagattaaacgtcacgaaccggtgggacgaatccaccgttggccact 120
Db 61 tccaaagggtccgagagattaaacgtcacgaaccggtgggacgaatccaccgttggccact 120

QY 121 gatgttcacgtggccaaacgcgcgcgatgtcgcacagtgtagccgcttcggtgcaggcg 180
Db 121 gatgttcacgtggccaaacgcgcgcgatgtcgcacagtgtagccgcttcggtgcaggcg 180

QY 181 gtcaaaaaggccccatggaagaagttcacaggtgcacaacgcgcggcgtgcatgcttaag 240
Db 181 gtcaaaaaggccccatggaagaagttcacaggtgcacaacgcgcggcgtgcatgcttaag 240

QY 241 ttcgcggacctcgcgcgagaagaacgcgcgagaagctcgtctgtagtgctgcccacc 300
Db 241 ttcgcggacctcgcgcgagaagaacgcgcgagaagctcgtctgtagtgctgcccacc 300

QY 301 ggtagaccggtgtcgatgacactcatttcgacattccaaacatggtctccgtgtttcgc 360
Db 301 ggtagaccggtgtcgatgacactcatttcgacattccaaacatggtctccgtgtttcgc 360
```

QY 361 tactatgcaggctggccgacaaagatcgccggaaagacctttcccgaggacaaacggcaag 420
Db 361 tactatgcaggctggccgacaaagatcgccggaaagacctttcccgaggacaaacggcaag 420

QY 421 ccgaattggcgttacgagccgatgggggtgtgtgctggtattgccagctggaacgcgact 480
Db 421 ccgaattggcgttacgagccgatgggggtgtgtgctggtattgccagctggaacgcgact 480

QY 481 ttctttacgtcggctggaagatagccccgcctcgccgcggcgtgctccttcattcttc 540
Db 481 ttctttacgtcggctggaagatagccccgcctcgccgcggcgtgctccttcattcttc 540

QY 541 aaagcctcggagaaaatccccgcgtggcggttcttggccctcgctcctctcttcgcagaagcc 600
Db 541 aaagcctcggagaaaatccccgcgtggcggttcttggccctcgctcctctcttcgcagaagcc 600

QY 601 ggattccctcctggagtcgtgcagttcctcactggagcacgagtgcgggtgaagcattg 660
Db 601 ggattccctcctggagtcgtgcagttcctcactggagcacgagtgcgggtgaagcattg 660

QY 661 gcgtgcacatggacattgcgaagatcagcttcacaaagatctgtcggcggtggccgcgc 720
Db 661 gcgtgcacatggacattgcgaagatcagcttcacaaagatctgtcggcggtggccgcgc 720

QY 721 gtcaagcaagcaacactcaagtccaacatgaagcgcgtcactctagaactgggggaaaaag 780
Db 721 gtcaagcaagcaacactcaagtccaacatgaagcgcgtcactctagaactgggggaaaaag 780

QY 781 ccaaccatcgtcttcaacgaagctcctctcgaacggcagtcgggggaatcggcgaaggat 840
Db 781 ccaaccatcgtcttcaacgaagctcctctcgaacggcagtcgggggaatcggcgaaggat 840

QY 841 ttctcaaaattcgggcaaaatttgggtccccccccctcctcgttttgcagtgcaaatgggaaat 900
Db 841 ttctcaaaattcgggcaaaatttgggtccccccccctcctcgttttgcagtgcaaatgggaaat 900

QY 901 ttagcggagaaaattccatggagtcocgtcatgggtcatttggaggctgtcagagatggcctt 960
Db 901 ttagcggagaaaattccatggagtcocgtcatgggtcatttggaggctgtcagagatggcctt 960

QY 961 ggcagaaacccattggaaacccaagaggagacgcatgggtcccttcgtcgaacaagtccagtagc 1020
Db 961 ggcagaaacccattggaaacccaagaggagacgcatgggtcccttcgtcgaacaagtccagtagc 1020

QY 1021 gacagagctcttgggttaacattgacgtttggaaggataccgcgcagctcctcactggcgctt 1080
Db 1021 gacagagctcttgggttaacattgacgtttggaaggataccgcgcagctcctcactggcgctt 1080

QY 1081 ggtagaaaggcgacaaaggattcgcgattgaaccgcagcatatttgtcaatcccaaacca 1140
Db 1081 ggtagaaaggcgacaaaggattcgcgattgaaccgcagcatatttgtcaatcccaaacca 1140

QY 1141 ggcagcaaaatttgggttgaggagatctttggcccccgcttctgtccattaaagacgttcaag 1200
Db 1141 ggcagcaaaatttgggttgaggagatctttggcccccgcttctgtccattaaagacgttcaag 1200

QY 1201 acggaagaagagccattgagattgccaatgacacgacttatgggctagcctcggtcatt 1260
Db 1201 acggaagaagagccattgagattgccaatgacacgacttatgggctagcctcggtcatt 1260

QY 1261 tataccaaatctctcaacaggggtctccggtgtctcgtcggcgtcgagacccggtggcgctc 1320
Db 1261 tataccaaatctctcaacaggggtctccggtgtctcgtcggcgtcgagacccggtggcgctc 1320

QY 1321 tcgatcaacttccccctttatcccccgagacacaaaactccgttttggcggtcatgaaacaatcg 1380
Db 1321 tcgatcaacttccccctttatcccccgagacacaaaactccgttttggcggtcatgaaacaatcg 1380

QY 1381 ggctcaggcagagagctaggcgaagaagggtcctcaaggcgtacttggagcccaagaccatt 1440
Db 1381 ggctcaggcagagagctaggcgaagaagggtcctcaaggcgtacttggagcccaagaccatt 1440

QY 1441 aatatccacgtcaacatagagtga 1464
Db 1441 aatatccacgtcaacatagagtga 1464

RESULT 6
US-09-677-682-4
; Sequence 4, Application US/09677682
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jon
; TITLE OF INVENTION: Compositions and Methods for Fummonisin Detoxification
; FILE REFERENCE: 35718/204101 (5718-111B)
; CURRENT APPLICATION NUMBER: US/09/677,682
; CURRENT FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1464
; TYPE: DNA
; ORGANISM: Exophiala spinifera
; FEATURE:
; OTHER INFORMATION: aldehyde dehydrogenase , fully spliced cDNA
US-09-677-682-4

Query Match 100.0%; Score 1464; DB 26; Length 1464;
Best Local Similarity 100.0%; Pred. NO. 0;
Matches 1464; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggttcttctgcctgacgaatacaagagtgaactcttcatcaacaatgaattcgtctcc 60
Db 1 atggttcttctgcctgacgaatacaagagtgaactcttcatcaacaatgaattcgtctcc 60

QY 61 tccaaaggggtccgagagattaaacgctcacgaacccgtgggacgaatccaccgttgcact 120
Db 61 tccaaaggggtccgagagattaaacgctcacgaacccgtgggacgaatccaccgttgcact 120

QY 121 gatgttcacgtggccaaacgcggccgatgtcgacagatgcagtagccgcttcggtgcaggcg 180
Db 121 gatgttcacgtggccaaacgcggccgatgtcgacagatgcagtagccgcttcggtgcaggcg 180

QY 181 gtcaaaaaggggcccatgggaagaagttcacaggtgcacaacgcgcggcgtgcatgcttaag 240
Db 181 gtcaaaaaggggcccatgggaagaagttcacaggtgcacaacgcgcggcgtgcatgcttaag 240

QY 241 ttccgggacctcgcgcgagaaagaaacgcgcgagaagctcgtcgtgtggagtgcgtgccacc 300
Db 241 ttccgggacctcgcgcgagaaagaaacgcgcgagaagctcgtcgtgtggagtgcgtgccacc 300

QY 301 ggtagaccgggtgctcgatgatcactcatttcgacatttccaaaacatggtctcogtgttccgc 360
Db 301 ggtagaccgggtgctcgatgatcactcatttcgacatttccaaaacatggtctcogtgttccgc 360

QY 361 tactatgcaggctggccgacaaagatcgccggaaagaccttcccgaggacaaacggcaag 420
Db 361 tactatgcaggctggccgacaaagatcgccggaaagaccttcccgaggacaaacggcaag 420

QY 421 ccgaattggcgttacgagccgatgggggtgtgtgctggtattgccagctggaacgcgact 480
Db 421 ccgaattggcgttacgagccgatgggggtgtgtgctggtattgccagctggaacgcgact 480

QY 481 tttctttacgtcggctgggaagatagcccccgccctcgccgcggcgtgctccttcattcttc 540
Db 481 tttctttacgtcggctgggaagatagcccccgccctcgccgcggcgtgctccttcattcttc 540

QY 541 aaagcctcggagaaaatccccgcgtggcggttcttggccctcgctcctctcttcgcagaagcc 600
Db 541 aaagcctcggagaaaatccccgcgtggcggttcttggccctcgctcctctcttcgcagaagcc 600

QY 601 ggattccctcctggagtcgtgcagttcctcactggagcacgagtgcagggtgaagcattg 660
Db 601 ggattccctcctggagtcgtgcagttcctcactggagcacgagtgcagggtgaagcattg 660

Db	601	ggattccctcctggagtcgtgcagttcctcactcgtgagcacgagtgacgggtggaagcattg	660
QY	661	gcgtcgcacatggacattgcgaagatcagcttcacaagatctgtcgcggctggccgcgcgc	720
Db	661	gcgtcgcacatggacattgcgaagatcagcttcacaagatctgtcgcggctggccgcgcgc	720
QY	721	gtcaagcaagcaacactcaagtcacaacatgaagcgcgtcactctagaaactcggggaaaaag	780
Db	721	gtcaagcaagcaacactcaagtcacaacatgaagcgcgtcactctagaaactcggggaaaaag	780
QY	781	ccaaccatcgtcttcaacgaagctcctctcgaacggcagtcgggggaatcggcaaaaagat	840
Db	781	ccaaccatcgtcttcaacgaagctcctctcgaacggcagtcgggggaatcggcaaaaagat	840
QY	841	ttctcaaaaattcggggcaaaatttgggtcccccctcctgttttgcctagtgcaatcggggaaaat	900
Db	841	ttctcaaaaattcggggcaaaatttgggtcccccctcctgttttgcctagtgcaatcggggaaaat	900
QY	901	ttagcggagaaaattccatggagtcctgcgtcatggctcaatttggaggctgtcagagatggcctt	960
Db	901	ttagcggagaaaattccatggagtcctgcgtcatggctcaatttggaggctgtcagagatggcctt	960
QY	961	ggccagaacccattggaaacccaagaggacgcatggtcccttcgtcgcacaagtcccaagtac	1020
Db	961	ggccagaacccattggaaacccaagaggacgcatggtcccttcgtcgcacaagtcccaagtac	1020
QY	1021	gacagagctcttgggtaacattgacgtttggcaaggataccgcgcagctcctcactggcgctt	1080
Db	1021	gacagagctcttgggtaacattgacgtttggcaaggataccgcgcagctcctcactggcgctt	1080
QY	1081	ggtagaaaggggcacaagggattcgcgattgaaaccgacgatatttgtcattccccaaacca	1140
Db	1081	ggtagaaaggggcacaagggattcgcgattgaaaccgacgatatttgtcattccccaaacca	1140
QY	1141	ggcagcaaaaatttgggttgaggagatcttttggccccgtcttgtccattaaagcgttcaag	1200
Db	1141	ggcagcaaaaatttgggttgaggagatcttttggccccgtcttgtccattaaagcgttcaag	1200
QY	1201	acggaagaagaggccattgagattgccaatgacacgacttatgggctagcctcggtcatt	1260
Db	1201	acggaagaagaggccattgagattgccaatgacacgacttatgggctagcctcggtcatt	1260
QY	1261	tataccaaaatctctcaacaggggtctcctgtgtctcgtcgcgtcgcagacgggtggcgctc	1320
Db	1261	tataccaaaatctctcaacaggggtctcctgtgtctcgtcgcgtcgcagacgggtggcgctc	1320
QY	1321	tcgatcaacttccccctttatccccgagacaaaaactccgttttggcggcatgaaaaaatcg	1380
Db	1321	tcgatcaacttccccctttatccccgagacaaaaactccgttttggcggcatgaaaaaatcg	1380
QY	1381	ggctcaggcagagagctaggcgaaagaagggtcctcaaggcgtacttggagccccagaccatt	1440
Db	1381	ggctcaggcagagagctaggcgaaagaagggtcctcaaggcgtacttggagccccagaccatt	1440
QY	1441	aatatccacgtcaacatagagtga	1464
Db	1441	aatatccacgtcaacatagagtga	1464

```

; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1464
; TYPE: DNA
; ORGANISM: Exophiala spinifera
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: aldehyde dehydrogenas
; US-09-677-682A-4

```

Query Match 100.0%; Score 1464; DB 26; Length 1464;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1464; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	atggttctttcgcctgacgaatacaagagtgaactcttcatcaacaatgaattcgtctcc	60
Db	1		
QY	61	tccaaggggtccgagagattaaacgtcacgaaccggtgggacgaatccacggttgccact	120
Db	61		
QY	121	gatgttcacgtggcccaacgcggccgatgtcgacagtgacgtagccgcttccggtgcaggcg	180
Db	121		
QY	181	gtcaaaaaagggcccatggaagaagtccacaggtgcacaacgcgcggcgtgcattgcttaag	240
Db	181		
QY	241	ttcgcggacctgcgcgagaagaacgcgcgagaagctcgtctgtggagtgcgtgccacc	300
Db	241		
QY	301	ggtagaccggtgtcgatgatactatttcgacattccaaacatggtctcgtgtttcgc	360
Db	301		
QY	361	tactatgcaggctggccgacaagatgcgcggaaagacctttcccgaggacaacggcaag	420
Db	361		
QY	421	ccgaattggcgttacgagccgatgggggtgtgtgctggtattgccagctggaacgcgact	480
Db	421		
QY	481	ttctttacgtcggctggaagatagccccgcctcgcgcgcggctgctcttcattcttc	540
Db	481		
QY	541	aaagcctcggagaaaatccccgcctgggcgttcttgggcctcgtctcttcgcagaaagcc	600
Db	541		
QY	601	ggattccctcctggagtcgtgcagttcctcaactggagcacgagtgacgggtgaagcattg	660
Db	601		
QY	661	gcgtcgacatggacattgcgaagatcagcttcacaagatctgtcggcgttgccgcgcgc	720
Db	661		
QY	721	gtcaagcaagcaacactcaagtccaacatgaagcgcgtcactctagaactgggggaaag	780
Db	721		
QY	781	ccaaccatcgtcttccaacgaagctcctctcgaacggcagtcggggggaatcggcaaggat	840
Db	781		

RESULT 7
US-09-677-682A-4
; Sequence 4, Application US/09677682A
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; TITLE OF INVENTION: Detoxification
; FILE REFERENCE: 35718/204101
; CURRENT APPLICATION NUMBER: US/09/677, 682A
; CURRENT FILING DATE: 2000-10-02

Db 781 ccaaccatcgtcttcaacgaagctcctctcgaacggcagtcgggggaatcggcaaaaggat 840
QY 841 ttctcaaaattcgggcaaaatttgggtcccccctcctcgtgttgcgtagtgcgaatgggaaaa 900
Db 841 ttctcaaaattcgggcaaaatttgggtcccccctcctcgtgttgcgtagtgcgaatgggaaaa 900
QY 901 ttagcggagaaaattccatggagtcgcgtcatggtcatttggagggtgtcagagatggcctt 960
Db 901 ttagcggagaaaattccatggagtcgcgtcatggtcatttggagggtgtcagagatggcctt 960
QY 961 ggccagaaccatttgaacccaagaggaacgcgtggtcccttcgtcgacaagtcccagtac 1020
Db 961 ggccagaaccatttgaacccaagaggaacgcgtggtcccttcgtcgacaagtcccagtac 1020
QY 1021 gacagagtcttgggttaacattgacgttggcaaggataccggcagctcctcactggcgctt 1080
Db 1021 gacagagtcttgggttaacattgacgttggcaaggataccggcagctcctcactggcgctt 1080
QY 1081 ggtagaaaaggcgacaaaggattcgcgattgaaacccgacgatatattgtcaatcccaaaacca 1140
Db 1081 ggtagaaaaggcgacaaaggattcgcgattgaaacccgacgatatattgtcaatcccaaaacca 1140
QY 1141 ggcagcaaaatttgggttaggagatcttttggccccctgttgcatttaagacgttcaag 1200
Db 1141 ggcagcaaaatttgggttaggagatcttttggccccctgttgcatttaagacgttcaag 1200
QY 1201 acggaagaagaggccattgagattgccaatgacacgaacttatgggctagcctcgggtcatt 1260
Db 1201 acggaagaagaggccattgagattgccaatgacacgaacttatgggctagcctcgggtcatt 1260
QY 1261 tataccaaatctcctaacacaggggtctccgtgtctcgtgcgcgtcgagaccgggtggcgctc 1320
Db 1261 tataccaaatctcctaacacaggggtctccgtgtctcgtgcgcgtcgagaccgggtggcgctc 1320
QY 1321 tcgatacaacttccctttatccccgagacacaaactccgttttggcgcatgaaacaatcg 1380
Db 1321 tcgatacaacttccctttatccccgagacacaaactccgttttggcgcatgaaacaatcg 1380
QY 1381 ggctcaggcagagagctaggcggaagggtcctcagggcgtacttggagccccagaccatt 1440
Db 1381 ggctcaggcagagagctaggcggaagggtcctcagggcgtacttggagccccagaccatt 1440
QY 1441 aatatccacgtcaacatagatga 1464
Db 1441 aatatccacgtcaacatagatga 1464

RESULT 8
US-09-677-682B-4
; Sequence 4, Application US/09677682B
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; TITLE OF INVENTION: Detoxification
; FILE REFERENCE: 35718/204101
; CURRENT APPLICATION NUMBER: US/09/677,682B
; CURRENT FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1464
; TYPE: DNA
; ORGANISM: Exophiala spinifera
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: aldehyde dehydrogenase, fully spliced cDNA

US-09-677-682B-4

Query Match 100.0%; Score 1464; DB 26; Length 1464;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1464; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 atggttcttctgcctgacgaatacaagaagtgaactcttctcatcaacaatgaattcgtctcc 60
Db 1 atggttcttctgcctgacgaatacaagaagtgaactcttctcatcaacaatgaattcgtctcc 60
QY 61 tccaaggggtccgagagatttaacgctcacgaaccctggggacgaatccaccgttgcact 120
Db 61 tccaaggggtccgagagatttaacgctcacgaaccctggggacgaatccaccgttgcact 120
QY 121 gatgttcacgtggcccaacgcgcgcgatgtcgacagtgacgtagccgttccgagggcg 180
Db 121 gatgttcacgtggcccaacgcgcgcgatgtcgacagtgacgtagccgttccgagggcg 180
QY 181 gtcaaaaaaggccccatgggaagaagttcacagggtgcacaacgcgcgcgtgcatgcttaag 240
Db 181 gtcaaaaaaggccccatgggaagaagttcacagggtgcacaacgcgcgcgtgcatgcttaag 240
QY 241 ttgcgcggacctgc 300
Db 241 ttgcgcggacctgc 300
QY 301 ggtagaccggtgtcgtgatgatcactcatttcgcacattccaaaacatggtctccgtgttccgc 360
Db 301 ggtagaccggtgtcgtgatgatcactcatttcgcacattccaaaacatggtctccgtgttccgc 360
QY 361 tactatgcagggtggccgacaagatcgccgggaagacaccttcccaggacacaacggcaag 420
Db 361 tactatgcagggtggccgacaagatcgccgggaagacaccttcccaggacacaacggcaag 420
QY 421 ccgaattggcggttacgagccgatgggggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 480
Db 421 ccgaattggcggttacgagccgatgggggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 480
QY 481 ttcttttacgctcggtggaagatagccccccgcctcgcgcgcgcgcgcgcgcgcgcgcgcgc 540
Db 481 ttcttttacgctcggtggaagatagccccccgcctcgcgcgcgcgcgcgcgcgcgcgcgcgc 540
QY 541 aaagcctcggagaaaatccccgcgtggcggttcttggcctcgtcctcctctcttcgcagaagcc 600
Db 541 aaagcctcggagaaaatccccgcgtggcggttcttggcctcgtcctcctctcttcgcagaagcc 600
QY 601 ggattccctcctcgtgagtcgtgcagttcctcactggagcacgagtgacgggtgaaagcattg 660
Db 601 ggattccctcctcgtgagtcgtgcagttcctcactggagcacgagtgacgggtgaaagcattg 660
QY 661 gcgtgcacatggacattgcgaagatcagcttcacaagatctgtcggcggtggccgcgcgc 720
Db 661 gcgtgcacatggacattgcgaagatcagcttcacaagatctgtcggcggtggccgcgcgc 720
QY 721 gtcaagcaagcaacacactcaagtccaacatgaagcgcgtcactctagaactggggaaaaag 780
Db 721 gtcaagcaagcaacacactcaagtccaacatgaagcgcgtcactctagaactggggaaaaag 780
QY 781 ccaaccatcgtcttcaacgaagctcctcctcgaacggcagtcgggggaatcgcaaaaggat 840
Db 781 ccaaccatcgtcttcaacgaagctcctcctcgaacggcagtcgggggaatcgcaaaaggat 840
QY 841 ttctcaaaaattcggggcaaaatttgggtccccccctcctcgttttgcctagtgcaaatggggaaaat 900
Db 841 ttctcaaaaattcggggcaaaatttgggtccccccctcctcgttttgcctagtgcaaatggggaaaat 900
QY 901 ttagcggagaaaattccatggagtcgcgtcatggtcatttggagggtgtcagagatggcctt 960
Db 901 ttagcggagaaaattccatggagtcgcgtcatggtcatttggagggtgtcagagatggcctt 960
QY 961 ggccagaaccatttgaacccaagaggaacgcgtggtcccttcgtcgacaagtcccagtac 1020


```
Db 961 ggcagaaacccattggaacccaagaggaacgcagtggtccctctgtcgacaagtcccaagtac 1020
QY 1021 gacagagtcttggttaacattgacgttggaaggatataccggcagctcctcactggcggtt 1080
Db 1021 gacagagtcttggttaacattgacgttggaaggatataccggcagctcctcactggcggtt 1080
QY 1081 ggtagaaaggcgacaaaggattcgcgattgaacccgacgatatattgtcaatcccaaacca 1140
Db 1081 ggtagaaaggcgacaaaggattcgcgattgaacccgacgatatattgtcaatcccaaacca 1140
QY 1141 ggcagcaaaatttggttgaggagatcttttgccccctgttgcatttaagacgttcaag 1200
Db 1141 ggcagcaaaatttggttgaggagatcttttgccccctgttgcatttaagacgttcaag 1200
QY 1201 acggaagaaggccattgagattgccaatgacacgacttatggcctagcctcggtcatt 1260
Db 1201 acggaagaaggccattgagattgccaatgacacgacttatggcctagcctcggtcatt 1260
QY 1261 tatacaaattctcaacagggtctcgtgtctcgtggtcgtcgtcgtcgtcgtcgtcgtcgtc 1320
Db 1261 tatacaaattctcaacagggtctcgtgtctcgtggtcgtcgtcgtcgtcgtcgtcgtcgtc 1320
QY 1321 tcgatcaacttccctttatccccgagacacaaactccgtttggcggcatgaaacaatcg 1380
Db 1321 tcgatcaacttccctttatccccgagacacaaactccgtttggcggcatgaaacaatcg 1380
QY 1381 ggctcaggcagagactagggcgaagggtcctcaaggcgtacttgagccccaaagaccatt 1440
Db 1381 ggctcaggcagagactagggcgaagggtcctcaaggcgtacttgagccccaaagaccatt 1440
QY 1441 aatatccacgtcaacatagagtga 1464
Db 1441 aatatccacgtcaacatagagtga 1464

RESULT 9
US-09-882-694-4
; Sequence 4, Application US/09882694
; GENERAL INFORMATION:
; APPLICANT: Duwick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; TITLE OF INVENTION: Detoxification
; FILE REFERENCE: 35718/208255
; CURRENT APPLICATION NUMBER: US/09/882,694
; CURRENT FILING DATE: 2001-06-12
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1464
; TYPE: DNA
; ORGANISM: Exophiala spinifera
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: aldehyde dehydrogenase, fully spliced cDNA
US-09-882-694-4

Query Match 100.0%; Score 1464; DB 33; Length 1464;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1464; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggtttcttcgctgacgaataacaaagagtgaactcttcatcaacaacaatgaattcgtctcc 60
Db 1 atggtttcttcgctgacgaataacaaagagtgaactcttcatcaacaacaatgaattcgtctcc 60
QY 61 tcaagggttccgagagattaacgctcaagaacccgtggagcaatccaccgttgccact 120
```

```
Db 61 tcaagggttccgagagattaacgctcaagaacccgtggagcaatccaccgttgccact 120
QY 121 gatgttcacgttgccaaacggtccgattgtcgacagtgacgtagccgtcttcggtgcaggcg 180
Db 121 gatgttcacgttgccaaacggtccgattgtcgacagtgacgtagccgtcttcggtgcaggcg 180
QY 181 gtcaaaaagggtcccatggaagaagtccaaggtgcacaacgcgcggtgcgtgcatgcttaag 240
Db 181 gtcaaaaagggtcccatggaagaagtccaaggtgcacaacgcgcggtgcgtgcatgcttaag 240
QY 241 ttccggtgacctcgcgagaagaacgcgcgagaagctcgtcgtctgtgagtgctgcccacc 300
Db 241 ttccggtgacctcgcgagaagaacgcgcgagaagctcgtcgtctgtgagtgctgcccacc 300
QY 301 ggtagaccgtgtcgtatgatcactcatttcgacatttccaaacatggtctcgtgtttcgc 360
Db 301 ggtagaccgtgtcgtatgatcactcatttcgacatttccaaacatggtctcgtgtttcgc 360
QY 361 tactatgcaggttggtccgacaaagatgcgcgaaagaccttcccgaggacaacgcgaag 420
Db 361 tactatgcaggttggtccgacaaagatgcgcgaaagaccttcccgaggacaacgcgaag 420
QY 421 ccgaattggcgttacgagccgatgggggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 480
Db 421 ccgaattggcgttacgagccgatgggggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 480
QY 481 ttcttttacgtcgttggaagatagccccccgcctcgcgcgcgcgcgcgcgcgcgcgcgcgc 540
Db 481 ttcttttacgtcgttggaagatagccccccgcctcgcgcgcgcgcgcgcgcgcgcgcgcgc 540
QY 541 aaagcctcggagaaaatccccgctggcggttcttggcctcgtcgtcgtcgtcgtcgtcgtcgt 600
Db 541 aaagcctcggagaaaatccccgctggcggttcttggcctcgtcgtcgtcgtcgtcgtcgtcgt 600
QY 601 ggattccctccttgagtgctgcaagtctcactggagcacagtgacgggtgaagcattg 660
Db 601 ggattccctccttgagtgctgcaagtctcactggagcacagtgacgggtgaagcattg 660
QY 661 gcgtcgcacatggacattgcgaagatcagcttcaagaatctgtcggcggtggccgcgcgc 720
Db 661 gcgtcgcacatggacattgcgaagatcagcttcaagaatctgtcggcggtggccgcgcgc 720
QY 721 gtcaagcaagcaacactcaagtccaacatgaagcgcgtcactctagaacctgggggaaaaag 780
Db 721 gtcaagcaagcaacactcaagtccaacatgaagcgcgtcactctagaacctgggggaaaaag 780
QY 781 ccaaccatcgtcttcaacgaagctcctcgcgaacgcgcgtcgcgcgcgcgcgcgcgcgcgcgc 840
Db 781 ccaaccatcgtcttcaacgaagctcctcgcgaacgcgcgtcgcgcgcgcgcgcgcgcgcgcgc 840
QY 841 ttctcaaaattcgggcaaaatttggttccccccctcctgtttgttagtgcaatgggggaaat 900
Db 841 ttctcaaaattcgggcaaaatttggttccccccctcctgtttgttagtgcaatgggggaaat 900
QY 901 ttagcggagaaaattccatggagtcctcattggctcatttggaggtgtgcagagatggctt 960
Db 901 ttagcggagaaaattccatggagtcctcattggctcatttggaggtgtgcagagatggctt 960
QY 961 ggcagaaacccattggaacccaagaggaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1020
Db 961 ggcagaaacccattggaacccaagaggaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1020
QY 1021 gacagagtcttggttaacattgacgttgcaaggtataccggcagctcctcactggcggtt 1080
Db 1021 gacagagtcttggttaacattgacgttgcaaggtataccggcagctcctcactggcggtt 1080
QY 1081 ggtagaaaggcgacaaaggattcgcgattgaacccgacgatatattgtcaatcccaaacca 1140
Db 1081 ggtagaaaggcgacaaaggattcgcgattgaacccgacgatatattgtcaatcccaaacca 1140
QY 1141 ggcagcaaaatttggttgaggagatcttttgccccctgttgcatttaagacgttcaag 1200
Db 1141 ggcagcaaaatttggttgaggagatcttttgccccctgttgcatttaagacgttcaag 1200
```


Db 1321 tcgatcaacttccctttatcccccgagacacaaactccgtttggcgcatgaacacatcy 1380
Qy 1381 ggctcaggcagagctagcggaagaaggctcaaggcgtacttgagcccaagaccatt 1440
Db 1381 ggctcaggcagagctagcggaagaaggctcaaggcgtacttgagcccaagaccatt 1440
Qy 1441 aatatccacgtcaacatagagtga 1464
Db 1441 aatatccacgtcaacatagagtga 1464

RESULT 11
US-08-612-853-1
; Sequence 1, Application US/08612853
; GENERAL INFORMATION:
; APPLICANT: Achatz, Gernot
; APPLICANT: Oberkofler, Hannes
; APPLICANT: Simon, Birgit
; APPLICANT: Unger, Andrea
; APPLICANT: Lechenauer, Erich
; APPLICANT: Hirschwehr, Reinhold
; TITLE OF INVENTION: Recombinant Cladosporium Herbarum
; TITLE OF INVENTION: Allergens
; NUMBER OF SEQUENCES: 117
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/612,853
; FILING DATE: 26-FEB-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AT94/00120
; FILING DATE: 02-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Friebe, Thomas E
; REGISTRATION NUMBER: 29,258
; REFERENCE/DOCKET NUMBER: 6530-021-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1491 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Cladosporium herbarum
; DEVELOPMENTAL STAGE: Spores and vegetative hyphae
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 1...1488
; OTHER INFORMATION:
US-08-612-853-1

Query Match 18.3%; Score 267.6; DB 10; Length 1491;
Best Local Similarity 53.0%; Pred. No. 3.1e-63;
Matches 766; Conservative 0; Mismatches 654; Indels 26; Gaps 8;

Qy 17 acgaatacaagagtgaactcttcatcaacaatagaattcgtctctcccaagggtccgaga 76
Db 41 acgagcagccgacccggctcttcatcaacaacagagttcgtaaggcgcaagaagcaga 100
Qy 77 gattaaacgtcacgaacccgtgggacgaatccaccgttgcacgtgattcaccgtggcca 136
Db 101 ccttcgatgtcatcaacccctccgacga---gagcgtgatcacccaggtccacgagccca 157
Qy 137 acgcgccgatgtcgacagtgcagtagccggttcggtgcaggggtcaaaaaaggggcccat 196
Db 158 ccgagaaaggtatgcgacatcgccgtcgccggccgcccgaagcccttc---gagggctcat 214
Qy 197 ggaagaagttcacaggtgcacaacgcgcggcggtgcatgttaagtgcgggacctcgcg 256
Db 215 ggagactggagacaccccgaaacccgtggcaagctgctcaacaaccccgcccaacctgttcg 274
Qy 257 agaaacacgcccagagaagctcgtctgtgagtgctgcacacccggtagaccggttoga 316
Db 275 agaaacacactgacctcttgcgtgcccgttgagtcgctcgacaacgcaagggccactcca 334
Qy 317 tgatacactcatttcgacattccaaacatggtctcctcgtgttctgctactatgcaggtgg 376
Db 335 tg---gcaagggtgacatcagcatgtgcgtccggtcgctcagatactacggtgtggg 391
Qy 377 ccgacaagatcgcgggaagaccccttcccgagacaaacgcgaagccgaattggcgttac- 435
Db 392 cggacaagatcacccggcaaggtcatcgacactagcccccgcacactttcaactacgtcaaga 451
Qy 436 --gagccgatgggggtgtgtgtgtattgcagctggaacgcgactttcttttacgtcg 493
Db 452 aggagcccatgt 511
Qy 494 gctggaagatagccccgcgcctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 553
Db 512 catggaagatcgcccgccgcttgcgttcggtgacacactgctcgtcctgaagactgctgagc 571
Qy 554 aatccccgctggggcttctggggcctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 613
Db 572 agacccctcttgggt 631
Qy 614 gactcgtgcagttcctcactgagcagca---gtgacgggtgaagcattggcgcgcaca 670
Db 632 gtgtcatcaacgtcatctccggttgcggcaagggtggtgtgtgtgtgtgtgtgtgtgtgtgt 691
Qy 671 tggacattgcgaagatcagcttcacaagatctgtcggcggtgagcgcgcgcgcgcgcgcgcgc 730
Db 692 tggacgtcgacaagggtggccttcacccggttccacgcgtgtgcggccgcgcgcgcgcgcgcgc 751
Qy 731 caacactcaagtccaaatgaagc 790
Db 752 ctgctgctcttccaaacttgaagaagggtcacccctcgagctcggtggaaggtcaccaaca 811
Qy 791 tcttcaacgaagctcctcctcgaaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 850
Db 812 ttgtcttcgagagc 871
Qy 851 tcggggaatttgggtcccccctcctggttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 910
Db 872 tcaaccacggccagtcgt 924
Qy 911 aattccatggagtcctcgtcatttggag-gctgtcagagatggtgtgtgtgtgtgtgtgtgtgt 969
Db 925 atctacgacaagttcgtccagaaggttcaaggagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 984
Qy 970 ccattggaacccaagagagcagcatggtcccttcgtcgacaagtcctcagtcagacagatc 1029
Db 985 ccttcgcccgc 1044
Qy 1030 ttgggttaacattgacgttggcaaggatatacc---gcgcagctcctcactggtgtgtgtgtgt 1086
Db 1045 atggagtacatccagggccgcaaggacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1104


```

; NAME/KEY: allele
; LOCATION: (3960)
; OTHER INFORMATION: PS4: polymorphic base G or A
; NAME/KEY: allele
; LOCATION: (4218)
; OTHER INFORMATION: PS5: polymorphic base C or T
; NAME/KEY: allele
; LOCATION: (4230)
; OTHER INFORMATION: PS6: polymorphic base C or T
; NAME/KEY: allele
; LOCATION: (4292)
; OTHER INFORMATION: PS7: polymorphic base C or T
; NAME/KEY: allele
; LOCATION: (4311)
; OTHER INFORMATION: PS8: polymorphic base A or G
; NAME/KEY: allele
; LOCATION: (4355)
; OTHER INFORMATION: PS9: polymorphic base T or G
; NAME/KEY: allele
; LOCATION: (4561)
; OTHER INFORMATION: PS10: polymorphic base G or A
; NAME/KEY: allele
; LOCATION: (4562)
; OTHER INFORMATION: PS11: polymorphic base G or A
; NAME/KEY: allele
; LOCATION: (4792)
; OTHER INFORMATION: PS12: polymorphic base A or G
; NAME/KEY: allele
; LOCATION: (4861)
; OTHER INFORMATION: PS13: polymorphic base G or A
; NAME/KEY: allele
; LOCATION: (5444)
; OTHER INFORMATION: PS14: polymorphic base T or C
; NAME/KEY: allele
; LOCATION: (5448)
; OTHER INFORMATION: PS15: polymorphic base C or T
; NAME/KEY: allele
; LOCATION: (5505)
; OTHER INFORMATION: PS16: polymorphic base C or T
; NAME/KEY: allele
; LOCATION: (5645)
; OTHER INFORMATION: PS17: polymorphic base G or C
; NAME/KEY: allele
; LOCATION: (5715)
; OTHER INFORMATION: PS18: polymorphic base C or A
; PCT-US01-17253-89

Query Match      16.5%; Score 241.6; DB 1; Length 9589;
Best Local Similarity 49.4%; Pred. No. 9.4e-56;
Matches 719; Conservative 9; Mismatches 713; Indels 15; Gaps 4;

QY 5 ttcttgcgtgacgaatacaagagtgaaactcttcatcaacaataatgctctctcca 64
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4121 ttctgaaccagacatccccctacaaccagctgttcatcaacaataatggaagatgcag 4180

QY 65 aggggtccgagagattaaacgtcacgaacccgtgggagcgaatccaccgttgccactgatg 124
   | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4181 tcagcaagaagaccttcccgacggtcaacccctaccacggtgggaggtcatygggacgtgg 4240

QY 125 ttcacgtggccaacgcggccgagtgacagtgacagtgacagtcggttcggtgcaggcggtca 184
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4241 ctgaaggtgaccgggctgatgtggatcgggccgtgaaagcagccccgggaagcttccgcc 4300

QY 185 aaaagggcccatggaagaattcacaggtgcacaacgcggcggtgcgtgcttaagtctcg 244
   | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4301 tggggtccccrctggcgggatggatgcctctgagcggggccggtgctgaackcctgg 4360

QY 245 cggacctgcggagagaagacgcggagagctcgctctgtgagtcgctgccaccggta 304
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4361 cagacctagtggagcgggatcgagctctacttggcctcactcgagaccttggacaatggga 4420

QY 305 gaccggtgctgatgatcaactcatttcgacattccaaacatggtctcctcggttctcgctact 364
```

```

Db 4421 agcctttccaagagtctttacgccttgagacttgatgaggtcatcaaggtgtatcggtact 4480
QY 365 atgcaggtcggccgacaagatgcgcggaaagacacctttcccgaggacaacggaagccga 424
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4481 ttgctggtggtgctgacaagtggcatggcaagaccatccccatgcatgcccagcattct 4540
QY 425 att--ggcgttacgagcgtatgggggtgtgtgctggtattgccagctggaacgcgactt 481
   | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4541 gcttcaccggcatgagccrctgtgtctgtctgctgagatcatcccgaggaaacttcccc 4600
QY 482 ttcttacgtcggctggaagatagccccgcctcccgccgctgctgctctctcatcttca 541
   | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4601 tggtcattgcagggttggaaacttgcgcggcactgcacacaggaacactgtgttatga 4660
QY 542 aagcctcggagaaaatccccgctgggctgttctgggctcgtcctctcttcgcagaaagccg 601
   || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4661 argtggcagagcagacccccctctctgccccctgtatttggcctccctcatcaaggagcag 4720
QY 602 gattccctcctggagtcgtgcagttcctcactgg--agcacgagtgacgggtgaagcat 658
   || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4721 gctttccccctgggtgtgaacatcatcaacgggtatggcccaacagcaggtgcggcca 4780
QY 659 tggcgtcgcacatggacattgcgaagatcagcttcacaagatctgtcggcggtggccgcg 718
   || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4781 tcgcccagcacrtggatgttgacaaagtgtccttcaccggttccaccgaggtggccacc 4840
QY 719 ccgtcaagcaagcaacacactcaagtccaacatgaagcgcgtcactctagaaactggggaaa 778
   || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4841 tgatccagaagaagcagctgctggccgagctgacatgagcagcgtggagcagtgccacg 4900
QY 779 agccaaccatcgtcttcaacgaagctcctctcgaacgagtcgggggaatcggaaggg 838
   || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4901 agagccccagcagctgctgctggccgagctgacatgagcagcgtggagcagtgccacg 4960
QY 839 atttctcaaaaatttcgggcaaaatttgggtccccccccctcctgttctgtagtgcgaatgggaa 898
   | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4961 aagcctgttcttc-----aacatgggcccagtgctgtgctggtcctccgggaccttcg 5014
QY 899 atttagcggagaaaattccatggagtcctcgtcatggtcatttggaggctgtcagagatggc 958
   || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5015 tggagaagaatccatctacaatgagtttctcgagagacccgtggagaaagcaagcagagga 5074
QY 959 ttggccagaacccattgggaacccaagagagcagcagctggttcccttcgtcgacaagtcccgat 1018
   || ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5075 aagtggggaaacccctttgagctggacacccacccagcaagggttcaggtggacaagagcag 5134
QY 1019 acgacagagttcttgggttaacattgacgtttggc--aaggataccgcgcagctcctcactg 1075
   || ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5135 ttgaacgagtcctaggtacatccagcttggccagaagagggcgcaaaactcctctgtg 5194
QY 1076 gcgttggtagaaaagggcgacaagggatttcgcgattgaaccgacgatatattgtcaatccca 1135
   ||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5195 gcggagagcgttttcggggagcgtggtttcttctcatcaagcctactgtcttttggcgctgc 5254
QY 1136 aaccagggcagcaaaaatttggtttgaggagagatcttttggcccccgcttctgtccattaaagcgt 1195
   | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5255 aggatgacatgagaattggcaaaagaggagatcttttggccctgtgcagccccctgttcaagt 5314
QY 1196 tcaagaagggaagaagagggccattgagattgccaatgacacgacttatgggtagcctcgg 1255
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5315 tcaagaagattgaggaggtggtgtgagaggggccaacaacacacaggtatggcctggctgcgg 5374
QY 1256 tcatttataccaaaatctctcaacaggggtctcctcggtctcgtcgtcgccgtcgagaccgggtg 1315
   | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5375 ctgtgtcacccgggatcttgacaaggccatgtacttaccacaggcactccagggccggga 5434
QY 1316 gcgtctcgatcaacttccccctttatccccgagacacaaactccgttttggcggaatgaaac 1375
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5435 ccgtgtgggyaaayacctacaacatcgtcacctgcacacgccaatttggaggtttaaagg 5494
QY 1376 aatcgggtcaggcagagagctaggtcgaagaagggtctcaaggcgtacttggagcccaaga 1435
   |||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Db 5495 aatctggaayggaggagctgggtgaggatgggcttaaggcctacacagaggtaaaga 5554
QY 1436 ccattaatatccacgt 1451
| | | | |
Db 5555 cggtcaccatcaaggt 5570

Search completed: April 27, 2002, 01:33:51
Job time: 21995 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2002, 00:19:28 ; Search time 150.7 Seconds
(without alignments)
2386.247 Million cell updates/sec

Title: US-09-882-694A-4
Perfect score: 1464
Sequence: 1 atggttcttgcctgacga.....tccacgtcaacatagagtga 1464

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues
Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	141.4	9.7	1842	4 US-09-221-294-1	Sequence 1, Appli
2	138.6	9.5	1455	4 US-09-651-941-8	Sequence 8, Appli
3	138.6	9.5	1524	4 US-09-655-270A-8	Sequence 8, Appli
4	138.6	9.5	12508	4 US-09-655-270A-1	Sequence 1, Appli
5	138.6	9.5	12523	4 US-09-651-941-1	Sequence 1, Appli
6	128.8	8.8	9515	1 US-08-920-812-13	Sequence 13, Appl
7	128.8	8.8	9515	1 US-08-920-827-13	Sequence 13, Appl
8	128.8	8.8	9515	1 US-08-921-177-13	Sequence 13, Appl
9	128.8	8.8	9515	1 US-08-362-577C-13	Sequence 13, Appl
10	128.8	8.8	9515	2 US-08-920-828-13	Sequence 13, Appl
11	121.6	8.3	4403765	4 US-09-103-840A-2	Sequence 2, Appli
12	121.6	8.3	4411529	4 US-09-103-840A-1	Sequence 1, Appli
13	115.6	7.9	1497	1 US-08-513-841-4	Sequence 4, Appli
14	115.6	7.9	1497	2 US-08-696-834-4	Sequence 4, Appli
15	115.6	7.9	1497	2 US-08-942-673-4	Sequence 4, Appli
16	115.6	7.9	1497	4 US-09-118-317-4	Sequence 4, Appli
17	115.6	7.9	4624	2 US-08-696-834-5	Sequence 5, Appli
18	86.4	5.9	728	4 US-08-998-416-758	Sequence 758, App
19	75.8	5.2	727	4 US-08-998-416-738	Sequence 738, App
20	63.8	4.4	1608	6 5221737-6	Patent No. 5221737
21	60.8	4.2	662	1 US-08-322-742-16	Sequence 16, Appl
22	58.2	4.0	2400	4 US-08-930-001-1	Sequence 1, Appli
23	56.2	3.8	402	1 US-08-346-611-1	Sequence 1, Appli
24	56.2	3.8	402	2 US-08-794-494-1	Sequence 1, Appli
25	55.6	3.8	402	1 US-08-346-611-3	Sequence 3, Appli
26	55.6	3.8	402	2 US-08-794-494-3	Sequence 3, Appli
27	53.8	3.7	595	1 US-08-076-091C-5	Sequence 5, Appli

28	53.8	3.7	595	1 US-08-285-641-5	Sequence 5, Appli
29	37.2	2.5	68750	3 US-09-335-409-1	Sequence 1, Appli
30	37.2	2.5	68750	4 US-09-568-102-1	Sequence 1, Appli
31	37.2	2.5	68750	4 US-09-567-969-1	Sequence 1, Appli
32	37.2	2.5	68750	4 US-09-568-480-1	Sequence 1, Appli
33	37.2	2.5	68750	4 US-09-568-486-1	Sequence 1, Appli
34	37.2	2.5	68750	4 US-09-568-472-1	Sequence 1, Appli
35	37.2	2.5	71989	4 US-09-443-501A-2	Sequence 2, Appli
36	34.8	2.4	4931	4 US-08-726-320-2	Sequence 2, Appli
37	34.8	2.4	4931	4 US-09-208-716-2	Sequence 2, Appli
38	34.2	2.3	1680	4 US-09-603-185-1	Sequence 1, Appli
39	34	2.3	1715	4 US-08-757-230A-1	Sequence 1, Appli
40	34	2.3	1715	5 PCT-US95-02315-1	Sequence 1, Appli
41	34	2.3	3664	1 US-07-880-913-1	Sequence 1, Appli
42	33.6	2.3	613	4 US-09-227-357-99	Sequence 99, Appl
43	33.6	2.3	1251	2 US-08-851-088-5	Sequence 5, Appli
44	33.6	2.3	5535	1 US-08-089-755A-1	Sequence 1, Appli
45	33.6	2.3	5535	1 US-08-089-755A-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-09-221-294-1
; Sequence 1, Application US/09221294
; Patent No. 6268138
; GENERAL INFORMATION:
; APPLICANT: Riccardo Dalla-Favera and
; APPLICANT: Alessandro Massimo Gianni
; TITLE OF INVENTION: A Retroviral Vector Capable of Transducing the
; TITLE OF INVENTION: Aldehyde Dehydrogenase-1 Gene and Uses of Said
; TITLE OF INVENTION: Vector
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM 330 466 DX2
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/221,294
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 42990-B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-278-0400
; TELEFAX: 212-391-0525
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1842 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: N
; ANTI-SENSE: N
; FRAGMENT TYPE: N-terminal
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..1568
US-09-221-294-1

Query Match 9.7%; Score 141.4; DB 4; Length 1842;

```
Best Local Similarity 48.3%; Pred. No. 1.1e-32;
Matches 570; Conservative 0; Mismatches 576; Indels 33; Gaps 5;

QY 284 tggagtcgctgccaccgctagaccggtgctgatgacactcatttcgacattccaaaaca 343
    ||||| || || || || || || || || || || || || || || || || || ||
Db 375 TGGAGTCAATGAATGGTGGAAACTCTATTCCAATGCATATCTGAATGATTAGCAGGCT 434

QY 344 tggttccggtgttcgctactatgcaggctgggcccgaagatgcgggaagaaccttcc 403
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 435 GCATCAAAACATTGCGCTACTGTGCAGGTTGGGCTGACAAGATCCAGGCCAGGCCGTA 494

QY 404 c-----cgaagacaacggcgaagccgaattggcgcttacgagccgatgggggtgtg 454
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 495 CAATACCAATTGATGGAAATTTTTACATATACAAGACATGAACCTATTGGGGTATGTG 554

QY 455 ctggtattgccagctggaacgcgacttttctttacgtcggtctggaagatagcccccgc 514
    || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 555 GCCAAATCATTCCTTGGAAATTCCCGTTGGTTATGCTCATTTTGGAAAGATAGGGCCTGCAC 614

QY 515 tcgcgcgcgctgctccttcattcttcaaacgctcggaataccccgctggcggttctgg 574
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 615 TGAGCTGTGGAACACAGTGGTTGTCAAAACCAGCAGACGAGAAACTCCTCTCACGTGCTCTCC 674

QY 575 gctcgctcctctcttcgcagaagccggattccctcctcgagtcgtgcagttcctcactg 634
    || ||||| || || || || || || || || || || || || || || || || || ||
Db 675 ACGTGGCATCTTNAATAAAGAGCGCAGGGTTTCCTCTCGAGTAGTGAATATTGTTCCTG 734

QY 635 g---agcacgagtgcgggtgaagcattggcgctgcacatggacattgcgaagatcagct 691
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 735 GTTATGGGCTTACAGCAGGGCGAGCCATTTCCTCTCACATGGATATAGACAAAGTAGCCT 794

QY 692 tcacaagatctgtcggcggtggcgcgccgtcaagcaagcaacactcaagtccaacatga 751
    ||||| |||| | | |||| | | |||| | | |||| | | |||| | | ||||
Db 795 TCACAGGATCAACAGAGGTTGSCAAGTTGATCAAAAGAGTGCCGGGAAAGCAATCTGA 854

QY 752 agcgctcactctagaactggggaagcccaaccatcgtcttcaacgaagctcctctcg 811
    || || || || || || || || || || || || || || || || || || || || ||
Db 855 AGAGGCTGACCTGGAGCTTGAGGAAAGAGCCCTTGCAATTGTTAGCTGATGCCGACT 914

QY 812 aacggcagtcgggggaatcggcaaaa-----ggatttctcaaaaattcgggcaaatitggg 865
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 915 TGGACAATGCTGTGAATTTCACACCATGGGGTATTCTACCAACAGGCCAGGCTGTTGTA 974

QY 866 tccccccctcctgtttgctagtgcgaatggggaatttagcggagaaattccatggagtcc 925
    | || || || | | | || || || || || || || || || || || || || || ||
Db 975 TAGCCGCATCCAGGATTTTGTGGAAGAAATCAATTTATG-----ATGAGTTTG 1022

QY 926 gtcattggtcattttggaggctgtcagagatggccttggccagaacccatttgaaacccaaga 985
    || | | |||| | | |||| | | |||| | | |||| | | |||| | | ||||
Db 1023 TTCGAAGGAGTGTGTAGCGGGCTAAGAAGTATATCCTTGGAAATCCTCTGACCCCGAGAG 1082

QY 986 ggaagcatggtcccttcgtcgacaagtcccagtcacgacagagctcttgggtaacattgacg 1045
    || || || || | | |||| | | |||| | | |||| | | |||| | | ||||
Db 1083 TCACCTCAAGGCCCTCAGATTGACAAGGAACAATATGATAAAATACTTGACCTCAATTGAGA 1142

QY 1046 ttggcaag--gataccgcgcagctcctcactgctggttggtagaaaaggcgacaaggggat 1102
    || || || || | | || || || || || || || || || || || || || || ||
Db 1143 GTGGGAAGAAAGAGGGGCCCAACTGGAATGTGGAGGAGGCCCGTGGGGGAATAAAGSCT 1202

QY 1103 tcgcgattgaaccgacgatatattgtcaattccccaaacccaggcgagcaaaatttggttgagg 1162
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1203 ACTTTGTCCAGCCACAGTGTCTCTAATGTTACAGATGAGATCGGCATTGCCAAAGAGG 1262

QY 1163 agatctttggcccccgcttctccattaaagacgcttcaagacggaagagggccatttgaga 1222
    ||||| ||||| || || || || || || || || || || || || || || || ||
Db 1263 AGATTTTGGACCAGTGCAGCAAAATCATGAAGTTTAAATCTTTAGATGACCTGATCAAAA 1322

QY 1223 ttgccaatgacacgacttatgggctagcctcgttcatttataccaaatctctcaacaggg 1282
    || || || || || || || || || || || || || || || || || || || || ||
Db 1323 GAGCAAAACAATACTTTCTATGGCTTATCAGCAGGAGTGTTTACCAAGACATTGATAAAG 1382

QY 1283 gtctcgtgtctcgtgcgcgcgagaccggtggcgtctcgatcaacttccccttatcc 1342
    | ||||| || || || || || || || || || || || || || || || || ||
```

```
Db 1383 CCATAACAANTCTCTCTCTCTCTGCAGGCAGGAACAGTGTGGTGAANTTCTATGGCGTGG 1442

QY 1343 ccgagacacaaaactccgtttggcggcatgaacaataatcggtcctcaggcagagactagcg 1402
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1443 TAAGTGCCAGTGCCCTTTGGTGGATTCAAGATGTCTGGAATGGNAGAGAACTGGGAG 1502

QY 1403 aagaagggctcaaggcgctacttggagcccaagaccatta 1441
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1503 AGTACGGTTTCCATGAATATACAGAGGTCAAAACAGTCA 1541

RESULT 2
US-09-651-941-8
; Sequence 8, Application US/09651941
; Patent No. 6355470
; GENERAL INFORMATION:
; APPLICANT: ROUVIER, PIERRE E
; APPLICANT: WALTERS, DANA M
; APPLICANT: RAINER, RUSS
; TITLE OF INVENTION: Genes Encoding Picric Acid Degradation
; FILE REFERENCE: BC1022 US NA
; CURRENT APPLICATION NUMBER: US/09/651,941
; CURRENT FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: 60/152,545
; PRIOR FILING DATE: 1999-10-03
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 8
; LENGTH: 1455
; TYPE: DNA
; ORGANISM: Rhodococcus erythropolis HL PM-1
US-09-651-941-8

Query Match 9.5%; Score 138.6; DB 4; Length 1455;
Best Local Similarity 48.1%; Pred. No. 6.6e-32;
Matches 686; Conservative 0; Mismatches 704; Indels 35; Gaps 9;

QY 34 ctcttcacacaaatgaattcgtctcctccaaagggtccgagagagattaaacgctcacgaac 93
    ||| |||| | | | | | | | | | | | | | | | | | | | | | | |
Db 28 ctcttcacacgcgacacaaactgaaccccatcgtcgacgggggagaccttcgactcgatcaac 87

QY 94 ccgtgggacgaataccaocgttgccactgatgttcaactggccaacgcggccgactgctcag 153
    ||| |||| | | | | | | | | | | | | | | | | | | | | | | |
Db 88 ccggccgacgggtcgcaacctggcca--gcgtcgccgagggccacggccgagcgtcgcg 144

QY 154 agtcagtagccgcttcggtagcggcggtcaaaaaggcccatggaaagaaattcacaaggt 213
    |||| | | | | | | | | | | | | | | | | | | | | | | |
Db 145 cgtcgggtcgaagccgcgaaggcgcgggcca-----ggacgtggcagcgcgtcgccccg 198

QY 214 gcacaaacgcgcggcgtgcattgattcgcggacctcgcgagaaacgcgcgagaaag 273
    || | | | | | | | | | | | | | | | | | | | | | | | | |
Db 199 gccacgaaccccgctgatgttccgtacgcgcgcgctgatcgaagaaacaaagaccgag 258

QY 274 ctgcctcgtctggagtcgctccacccggttagaccgggtgctcgatgatcactcatttcgac 333
    |||| | | || || | | | | | | | | | | | | | | | | | | |
Db 259 ctgcgccagctgcagagtcggaacatgggcaagccccatccgcgagtcgctcgggactcgac 318

QY 334 attccaaaacatggctcctcgtttctactactatgcaggctggggccgacaagatcgccgga 393
    | | | | | |||| | | | | | | | | | | | | | | | | | | | |
Db 319 ctgcgcatcatgatcgagacgctcagtagtactcgcgggctcgtgaccaagatcgagggc 378

QY 394 aagacctttcccgaggacaacggcaagccgaattg---gcgttacgagccgatgggggtg 450
    | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 379 cgaacgacgcgcggccgcgcgcgtttctcactcaactacacctgcgtgagccgactcgtgtg 438

QY 451 tgtgctggtatttgcagctggaacgcgacttttcttactcgtggtcgtggaagatagcccc 510
    | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 439 gtgggcgcctcactcctcctggaatttctcgtcagtcgagcggtctggaagatcgcccc 498

QY 511 gcctcgcgcgcggctgctccttcacttcaaaagcctcggagaaatccccgctgggcgtt 570
    || || || || | | | | | | | | | | | | | | | | | | | | |
Db 499 gctcttgcgatgggcaacgccaatcgtgctgaagcctgcgcagctcgaccactcgtgccc 558
```


QY 1215 cattgagattgccaatgacacgacttatagggttagcctcggtcatttataccaaaatctct 1274
Db 5230 gataccctggccaacgacacacctcttcgggtgctggccggtctcttaccacgggatgt 5289
QY 1275 caacaggggtctcgtgctcgtcgccgctcgagaccggtgggtctcgatcaactccc 1334
Db 5290 gggcgcgactcggttcgagacgctcgacgccggcaacgtgtggatcaacagctg 5349
QY 1335 cttatccccgagacaaaactccgtttggcggtcatgaaacaatcgggtcagggcagaga 1394
Db 5350 gggagtgtcaaccgggtcgccgtatcgaggttcgggcagagcggctacggcagoga 5409
QY 1395 gctaggcgaagaaggctcaaggcgtacttggagcccaagaccat 1439
Db 5410 cctcggcaggcggtccatcgaagcttcaccaaggagagagcat 5454

RESULT 5

US-09-651-941-1
; Sequence 1, Application US/09651941
; Patent No. 6355470
; GENERAL INFORMATION:
; APPLICANT: ROUVIER, PIERRE E
; APPLICANT: WALTERS, DANA M
; APPLICANT: RAINER, RUSS
; TITLE OF INVENTION: Genes Encoding Picric Acid Degradation
; FILE REFERENCE: BC1022 US NA
; CURRENT APPLICATION NUMBER: US/09/651,941
; CURRENT FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: 60/152,545
; PRIOR FILING DATE: 1999-10-03
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 1
; LENGTH: 12523
; TYPE: DNA
; ORGANISM: Rhodococcus erythropolis HL PM-1
US-09-651-941-1

Query Match 9.5%; Score 138.6; DB 4; Length 12523;
Best Local Similarity 48.1%; Pred. No. 1.8e-31;
Matches 686; Conservative 0; Mismatches 704; Indels 35; Gaps 9;
QY 34 ccttcacatcaaatgaattcgtctcctccaaaggggtccgagagattaaacgctcacgaac 93
Db 4046 ctgcgtacatcggcacaaactgaccccatcgtcgacggggcgacacctcgactcgatcaac 4105
QY 94 ccgtgggacgaatccaccgttgccactgatgttcacgtggccaaacggtccggtcgtcgcac 153
Db 4106 ccggcgacgggtcgacacctggcca--ggtcgccgagggccacggccggtcggtcggtcgcg 4162
QY 154 agtgcagtagccgcttcggtgcagggcggtcaaaaagggtcccatggaagaagttcacaggt 213
Db 4163 cgtgcgtcgaaagccgcgaagggcggtccca-----ggaagtggaagcgtgcgtgcgtgcg 4216
QY 214 gcacaacgcggtggtgcgtgcatgcttaagtctcggtgacacctcgccgagagaacgcgagag 273
Db 4217 gccacggaacccgcctgatgttcgtacgcgcgctgatcgaggaacacaaagaccgag 4276
QY 274 ctgcgtcgtgagtcgctgctgccaccggttagacgggtgtcgatgatcactcatttcgac 333
Db 4277 ctgcgcagctgcagagtgcgggacatgggcaagcccatccgcgagtcgctcgggagtcgac 4336
QY 334 attccaaacatgggtctcgtgttcgtctactatgcaggtggccgacaaagatcgccgga 393
Db 4337 ctgcgatcatgatcgagacgctcgagtacttcggttcggttcgtgacaaagatcgagggc 4396
QY 394 aagacctttccgaggaacacggcaagccgaattg--gcgttacgagccggtgggtg 450
Db 4397 cgaacgacgcggcgcccggttccttcactacacccctcggtgagccggtggtg 4456

QY 451 tgtgtggtattgccagctggaaacgcgacttttttaccgtcggtcggtggaagatatagcccc 510
Db 4457 gtggcgccatcactccctggaatttctcgtcagtcagcggtctggaagatcgcccc 4516
QY 511 gccctgcgcgcgggtgctctccttcaatcttcaaaagcctcgagagaaaatcccgctggcggtt 570
Db 4517 gctcttgcatgggcaacgcacatcgtgctgaagcctcgcgagctcgacacactcgtgccc 4576
QY 571 ctgggctcgtcctctctcttcgcgagaagccgggattccctcctcgtggagtcgtgcagttcctc 630
Db 4577 gtggcaactcggcgagctcgccctcgagggggttcgcccgggttcgagcctcgtgcaacgtcctg 4636
QY 631 actggagcacga--gtgacgggtggaagcattggcgctgcacatggacattgcgaagatc 687
Db 4637 cccggcgcggtcggttagcggttaacgccttggtgtagcaccatcggtcggtcggaaggtg 4696
QY 688 agcttcacaagatctgtcgcggtggtggcgccgcgtcaagcaagcaacactcaagtccaac 747
Db 4697 acgttcaccgggtcgaccgaggtcgccagcagatcggtcggtggtggcgccgacggtcctc 4756
QY 748 atgaagcgtcactctagaaactgggggaaagcaaacacatcgtcttcaacgaag-ctcc 806
Db 4757 atcaag--gcttcgctggagctggcggaagctcgtcgtcgtcgttcggttcggtcgtcgtcgt 4813
QY 807 tctcgaacggcagtcgggggaatcggcaaa----ggatttctcaaaattcgggcaaat 861
Db 4814 tccccgaaggcgtcgacgcgtggtcttccagcgatgtacagcaacacggtgagacc 4873
QY 862 tgggtccccccctcctgtttgctagtgcattgggaaatttagcggagaaattccatgga 921
Db 4874 tgcacggcgccgagcaggttgcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 4929
QY 922 gtccgtcatgggtcatttggaggtcgtcagagatgggttggcagaacacattggaaccc 981
Db 4930 gctcgtccaggcacgtgtagagcggtgggtggcgacccggtgcgtcgtcgtcgtcgtcgtcgtcgt 4989
QY 982 aagaggagcagtggtcccttcg-----tcgacaagtcctcagtagcacagagtcgttggg 1034
Db 4990 gatcggtccgttgatcagtgccgagcagcgggtggtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 5049
QY 1035 taacattgacgttggcaaggataccgcgcagctcctcactgctcgttggtagaaaggcga 1094
Db 5050 gaccgaggaaggcgcacgctgatcagcgtggtggcagacagtcgccgacggagcgcgga 5109
QY 1095 caagggattcgcgattgaaccgacgatatatttgcatacccaaacacgaggaacaaatttg 1154
Db 5110 gcagggattctactaccgtccgacgctcttctcggaggtcaccgcggacatgcgcatcgc 5169
QY 1155 gtttgaggagatctttggccccgtcttgcattatgaagcgttcaagacggaagagagc 1214
Db 5170 tcgggagagatcttcggaccgtgctgctggtgctggtgctcggaggtcagaggagagagagc 5229
QY 1215 cattgagattgccaatgacacgacttatgggtagcctcgtcgtcatttataccaaatctct 1274
Db 5230 gatcaccttggtccaaacgacacacctcttcgggtggtggcggtcgtcttcaaccgcgcatgt 5289
QY 1275 caacaggggtctcgtgctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 1334
Db 5290 ggccgcgcactcgtggttcgcgacagcgtcgcacgcggcaacgtgtggtatcaacagctg 5349
QY 1335 ctttatccccgagacacaaactccgtttggcggtcatgaaacaacatcggtcgtcaggtcagaga 1394
Db 5350 gggagtgtcaaccggcggtcgccgtatcgaggttcgggcagagcgttcggtcgtcgtcgtcgtcgtcgt 5409
QY 1395 gctaggcgaagaagggtcctcaaggcgtacttggagcccaagaccat 1439
Db 5410 cctcggccaggcggtccatcgaagcttcaccaaggagagagcat 5454

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2002, 00:40:26 ; Search time 706.01 seconds
(without alignments)
3560.237 Million cell updates/sec

Title: US-09-882-694A-4
Perfect score: 1464
Sequence: 1 atggttcttgcctgacga.....tcacgtcaacatagagtga 1464

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1464	100.0	1464	22 AAD02692	Exophiala spinifer
2	267.6	18.3	1491	16 AAQ86278	Cladosporium herba
3	245.8	16.8	2022	21 AAF12996	Aspergillus oryzae
4	245.4	16.8	1710	21 AAF07491	Fusarium venenatum
5	239.2	16.3	1488	16 AAQ86275	Alternaria alternata
6	217	14.8	1980	23 ABL10879	Drosophila melanog
7	211.6	14.5	3442	23 AAS70764	DNA encoding novel
8	194.8	13.3	2700	7 AAN60401	Sequence constitut
9	191	13.0	3463	21 AAC78074	Human cancer assoc

10	190	13.0	1646	21	AAF15973	Human prostate can
11	184.6	12.6	2036	20	AAZ33590	Human breast tumou
12	182	12.4	1541	22	AAF82082	Human aldehyde deh
13	179.6	12.3	2071	21	AAF18274	Lung cancer associ
14	174.2	11.9	1772	21	AAZ56975	Arabidopsis aldehy
15	172.2	11.8	1863	21	AAZ56973	Arabidopsis aldehy
16	172.2	11.8	2182	23	ABL06647	Drosophila melanog
17	168.6	11.5	109519	22	AAS08693	Micromonospora DNA
18	161.6	11.0	6291	23	ABL06646	Drosophila melanog
19	156.6	10.7	6153	23	ABL10878	Drosophila melanog
20	154	10.5	1614	21	AAZ56974	Arabidopsis aldehy
21	153.6	10.5	1589	21	AAC44854	Arabidopsis thalia
22	141.4	9.7	1842	15	AAQ72450	Aldehyde-dehydroge
23	138.6	9.5	1524	21	AAAS3945	ORF4 sequence enco
24	138.6	9.5	12508	21	AAAS3941	12.5 kb picric aci
25	128.8	8.8	9515	15	AAQ55145	Pseudomonas aerugi
26	126	8.6	1473	23	AAS54367	Pseudomonas aerugi
27	121.6	8.3	4403765	22	AAI99683	Mycobacterium tube
28	121.6	8.3	4411529	22	AAI99682	Mycobacterium tube
29	118.2	8.1	1473	23	AAS52323	E. coli DNA for ce
30	117	8.0	1446	22	AAF61032	P. putida KT2440-a
31	117	8.0	1731	21	AAC50961	Arabidopsis thalia
32	117	8.0	1735	21	AAC44647	Arabidopsis thalia
33	115.6	7.9	1497	15	AAQ74471	L-sorbose dehydr
34	115.6	7.9	1497	16	AAT04279	Gluconobacter oxyd
35	115.4	7.9	2735	23	ABL03405	Drosophila melanog
36	114	7.8	3278	23	ABL24684	Drosophila melanog
37	113.8	7.8	1425	22	AAH22876	R. ruber cdd gene
38	113.8	7.8	10480	22	AAH22877	R. ruber 10kb gene
39	110.8	7.6	4624	16	AAT04318	Expression vector
40	107.6	7.3	1226	21	AAC33109	Arabidopsis thalia
41	106.2	7.3	1232	21	AAC50967	Arabidopsis thalia
42	105.6	7.2	715	22	AAS41088	cDNA encoding nove
43	104	7.1	1452	23	AAS51421	Pseudomonas aerugi
44	100.6	6.9	1827	21	AAC37522	Arabidopsis thalia
45	97.8	6.7	1446	23	AAS56094	Salmonella typhi D

ALIGNMENTS

RESULT 1
AAD02692
ID AAD02692 standard; DNA; 1464 BP.
XX
AC AAD02692;
XX
DT 02-MAY-2001 (first entry)
XX
DE Exophiala spinifera aldehyde dehydrogenase coding sequence.
XX
KW Fumonisin; degradative enzyme; transport enzyme; flavin monooxygenase;
KW aldehyde dehydrogenase; permease; p-glycoprotein; fumonisin esterase;
KW detoxification; mycotoxin; animal feed; human feed; silage;
KW transgenic plant; transgenic animal; microbial spray; ds.
XX
OS Exophiala spinifera.
XX
FH Key Location/Qualifiers
FT CDS 1..1464
FT /*tag= a
FT /product= "E. spinifera aldehyde dehydrogenase"
XX

PN WO200105980-A1.

XX 25-JAN-2001.

XX 14-JUL-1999; 99WO-US15824.

XX 14-JUL-1999; 99WO-US15824.

XX (PION-) PIONEER HI-BRED INT INC.

PA (CURA-) CURAGEN CORP.

AC	AQ86278;
XX	
XX	25-OCT-1995 (first entry)
XX	
DE	Cladosporium herbarum allergen Clah53 cDNA.
XX	
KW	Fungal spore; allergen; Clah53; allergy; aldehyde dehydrogenase; ds.
XX	
OS	Cladosporium herbarum.
XX	
FH	Key Location/Qualifiers
FT	CDS 1..1491
FT	/tag= a
FT	/product= Clah53
XX	
PN	WO9506121-A.
XX	
PD	02-MAR-1995.
XX	
PF	24-AUG-1994; 94WO-AF00120.
XX	
PR	27-AUG-1993; 93AT-0001725.
XX	
PA	(BIOM-) BIOMAY PRODN & HANDELSGES MBH.
XX	
PI	Achatz G, Breitenbach M, Ebner C, Hirschwehr R;
PI	Kraft D, Lechenauer E, Oberkofler H, Prillinger H;
PI	Simon B, Unger A;
XX	
DR	WPI; 1995-106850/14.
DR	P-PSDB; AAR71891.
XX	
PT	Allergens derived from Cladosporium herbarum spores - also
PT	recombinant DNA for expressing the allergens, useful for in vitro
PT	allergy detection
XX	
PS	Claim 1; Page 8-9; 35pp; German.
XX	
CC	Spores of Cladosporium herbarum are the most common fungal spores
CC	found in the air; they can cause allergic reactions. Various Clah
CC	allergens and sequences encoding them have now been isolated. The
CC	mature Clah53 allergen has mol. wt. 53 kD and is encoded by cDNA
CC	sequence AQ86278. The allergen has homology to aldehyde
CC	dehydrogenases. Potential epitopic subfragments were identified by
CC	computer analysis of the amino acid sequence. See AAR71892-R71906 for
CC	potential B-cell epitopes and AAR72615-R72627 for potential T-cell
CC	epitopes.
XX	
SQ	Sequence 1491 BP; 325 A; 474 C; 395 G; 297 T; 0 other;
Query Match 18.3%; Score 267.6; DB 16; Length 1491;	
Best Local Similarity 53.0%; Pred. No. 5.3e-70;	
Matches 766; Conservative 0; Mismatches 654; Indels 26; Gaps	
Qy	17 acgaatacaagagtgaactcttcatcaacaatgaattcgtctcctccaagggtccgaga 76
Dd	41 acgagcgcgcacggcctcttcatacaaacgagttcgtaagggccaaaggaaga 100
Qy	77 gattaacgctcacgaaccctgggacgaatccaccgttgccactgatgttcacgtggcca 136
Dd	101 ccttcgatgtcatcaaccccctccagca--gagcgtgatcaccagggtccacaggcca 157
Qy	137 acgcgccgatgtcgacagtgcagtaccgttcgttcggtgcaggcgggtcaaaaaggccccat 196
Dd	158 ccgagaaggatgtcgacatgcgcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 214
Qy	197 ggaagaagtccacaggtgcacacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 256
Dd	215 ggagactggagacacccgagaaacgctggcagctgctcaacaacctcgcccaacctgttcg 274
Qy	257 agaagaacgcgcgagaagctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 316

Db	275	agaaga	caactg	accctc	cttgctg	cgcgttg	agtcg	ctcg	acaa	cggca	ggcc	acttcca	334
QY	317	tgatc	actc	atttcg	acattc	ccaacat	gggtcc	cggtgt	ttcg	ctact	atgc	aggtggg	376
Db	335	tg--	-gca	gggtg	acatc	agcatg	tgcgt	ccggctg	cc	tcag	atac	agtggtggg	391
QY	377	ccgac	aatc	gcggga	agac	ctttcc	cgagg	acaa	cggca	agccga	attggc	gttac-	435
Db	392	cggac	aatc	acccg	caaggt	catcg	acact	acg	ccg	acact	ttca	actacg	451
QY	436	--	gagc	gatg	gggtgt	gtgtg	ctgt	attg	ccag	ctgga	acg	actttct	493
Db	452	aggag	cccat	gtgtg	ccggtc	agatc	attcc	ctgga	acttcc	cttcc	cttcat	gtggg	511
QY	494	gctgg	aatag	accc	ccgc	ccctcg	ccgc	cggtg	ctct	ctctc	atcttca	aaagc	553
Db	512	catgg	aatc	gcgg	cccat	gtctg	gggt	aaac	atg	ctg	ctc	tgag	571
QY	554	aatcc	ccgc	gtgg	cggtt	ctg	ggcc	ctg	ctct	ctctt	cgca	agccg	613
Db	572	agacc	ctt	gtgtg	ctc	gtc	gtc	gcag	ccctc	gtca	aggg	ccggtt	631
QY	614	gagtc	gtc	agttc	ctc	actg	gag	cacga--	-gtg	acgg	gtg	aagc	670
Db	632	gtg	catc	aacg	tcattt	ccggtt	cggca	aggtc	gtg	tcg	ctct	cttctc	691
QY	671	tggac	attg	cga	agatc	agctt	caca	agat	ctgt	gcgg	gtg	gccgc	730
Db	692	tggac	gtc	aca	aggtg	gcctt	cac	cggtt	ccac	cggtt	gtc	ggccg	751
QY	731	caac	actc	aatg	tcca	acatg	aa	gcgc	gtc	actc	t	agaa	790
Db	752	ctg	ctg	cc	cttcca	acttga	aga	aggtc	ac	ccctc	gag	tc	811
QY	791	tctt	caa	cga	agctc	ctctc	ga	acgg	cagtc	gggg	gaatc	ggca	850
Db	812	ttg	tctt	cga	ggac	gcg	atattg	acaa	cgc	atctc	atg	gttca	871
QY	851	t	cg	ggca	aat	ttgg	gttcccc	ctc	ctgtt	gtc	agtgc	aaatg	910
Db	872	tca	acc	acg	ccag	tgtgt	gtgtgt	gtgtgt	gtgtgt	gtgtgt	gtgtgt	gtgtgt	924
QY	911	aatt	ccat	ggag	tcg	tc	atggttc	atggttc	atggttc	atggttc	atggttc	atggttc	969
Db	925	atc	acg	aca	agttc	gtcc	aga	agttc	aa	ggag	gcgc	acaga	984
QY	970	ccatt	g	aa	ccca	agag	acg	catg	gtcc	ctt	gtc	gaca	1029
Db	985	ccct	t	gc	gcgc	acac	attc	ccag	gtc	ctc	aggtt	ccaa	1044
QY	1030	ttgg	g	ta	acattg	ac	gttgg	caagg	atacc--	-gcg	cag	ctc	1086
Db	1045	atg	gag	tac	atcc	agc	cgga	agc	gcgc	cggtg	ccac	ctc	1104
QY	1087	aagg	gcg	aca	agg	attc	gcg	attg	aa	ccg	acg	ata	1146
Db	1105	aagg	gtg	aca	agg	gtact	ctt	cattg	ag	cccc	accat	cttctc	1164
QY	1147	aaa	attg	gtt	gag	gag	atctt	tg	gg	cccc	ctt	gttcc	1206
Db	1165	aag	atc	gtg	aa	agg	gag	atctt	cg	cccc	ctc	gtc	1224
QY	1207	gaag	agg	ccatt	gag	attg	cca	atg	ac	ag	actt	atg	1266
Db	1225	gagg	acg	ccat	caag	ctc	ggca	acg	ccag	cac	ata	cggtt	1284
QY	1267	aaat	ctc	tca	ac	aggg	gttct	ccg	gtc	gtc	gtc	gtc	1326
Db	1285	aaga	ac	ctca	ac	acg	ccatc	gag	gttct	cca	acg	ctc	1344
QY	1327	aact	cccc	tttat	cccc	gag	acac	aa	actc	ogt	ttt	ggc	1386
Db	1345	aac	actt	aca	ac	cc	ctcc	acc	acag	atg	cc	ctt	1404

Db 576 acctctctccgcatacaattgcctgcaagctgataccaggagccggtttccaccagggtgt 635

QY 618 cgtgcagttcctcactgagcacgagtga--cgggtgaagcattggcgtcgacatgga 674

Db 636 catcaacgtcatcactggtttcggaagatcgccggtgctgccatgtccgctcacatgga 695

QY 675 cattggaagatcagcttcacaaagatctgtcggcgggtggccgcgcgtcaagcaagcaac 734

Db 696 cattgacaagattgcctttactggttcaaccgtttgcggcgtcaaatcatgaagtctgc 755

QY 735 actcaagtcacaacatgaagcgcgtcacctctagaactggggaaaaagccaacca-----t 788

Db 756 ggctggctccaacttgaaagaggtcacctcttgagctcgaggcaagagcccccaacattgt 815

QY 789 cgtctcaacgaagctcctctcgaacggcagtcgggggaatcggcaaggatttctcaaa 848

Db 816 cttcgcgcagcagatcttgacgaggctatccactgggtcaacttgggtatttacttcaa 875

QY 849 attcggggcaaatgttggttccccccctcctgtttgctagtgcaatggggaaaaatttagcgga 908

Db 876 ccacggacaggtctgttgctggttgcgtatctacgtccaagaagagatctacgacaa 935

QY 909 gaaattccatggagtc--cgtcatggctcatttgaggctgtgcagagatggcttggccag 966

Db 936 gttcatccagcgttcaaggagcggcgtgctcagaacgtgttggtagcccatcgccgc 995

QY 967 aacctattggaacccaagaggacgcgatggtccctctgctgcacaagtcccagtagcacaga 1026

Db 996 gacac-----tcagggtcctcaagtctcgcagctccagttcgaccgt 1038

QY 1027 gtcttggttaacattgacgttggcaagga---taccgcgcagctcctcactggcgttgg 1083

Db 1039 atcatgggtacatcgaggagggaagaagtctggcgcgaccatcgagactggtggcaac 1098

QY 1084 agaaaggcgacaaaggtatcgcgattgaaaccgacgatatttgtcaatcccaaacaccaggc 1143

Db 1099 cgttaagggtgacaagggttacttcatcgagcccaaatcttctccaactgaacaggagac 1158

QY 1144 agcaaaatttggttgaggagatcttggccccgtcttgtccattaaagacgttcaagacy 1203

Db 1159 atgaagattcagcaagaagagatcttcggccccgtctgcacaatctccaagttcaagaca 1218

QY 1204 gaagaagaggccattgagattgccaatgacacgacttatgggctagctcgtcatttat 1263

Db 1219 aaggcgcagctcatcaagattggcaacaacaccacatacgggtcttctggccgctgtacac 1278

QY 1264 accaaatctctcaacagggtctcctcgtgtctcgtcggcgtcgcagaccggtggcgtctcg 1323

Db 1279 acatccaacctcaccactgccactgcgaagtggccaacgcgtccgtgcaggaaactgtctgg 1338

QY 1324 atcaacttccccctttatccccgagacacaaaactcgtttggcggcatgaanaacaatcgggc 1383

Db 1339 gtcaactcctacaacactcttcaactggcagcttcccttcgagggttacaaggagtctggt 1398

QY 1384 tcaggcagagagctagggcgaagaagggtcctaaggcgtacttggagcccaagaccattaat 1443

Db 1399 attgggcgcgagttgggagaggcggcgctgggacaactacatccagaccgaacccgtgtct 1458

QY 1444 atcc 1447

Db 1459 attc 1462

RESULT 6

ABL10879

ID ABL10879 standard; cDNA; 1980 BP.

XX

AC ABL10879;

XX

DT 26-MAR-2002 (first entry)

XX

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 27119.

XX

KW Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

XX

OS Drosophila melanogaster.

XX

PN WO200171042-A2.

XX

PD 27-SEP-2001.

XX

PF 23-MAR-2001; 2001WO-US09231.

XX

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX

PA (PEKE) PE CORP NY.

XX

PI Venter JC, Adams M, Li PWD, Myers EW;

XX

DR WPI; 2001-656860/75.

DR P-PSDB; ABB66776.

XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions .

XX

PS Claim 1; SEQ ID NO 27119; 21pp + Sequence Listing; English.

XX

CC The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABL01840-ABL16175) and the encoded proteins

CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX

SQ Sequence 1980 BP; 471 A; 527 C; 540 G; 442 T; 0 other;

Query Match 14.8%; Score 217; DB 23; Length 1980;

Best Local Similarity 50.0%; Pred. No. 1.le-54;

Matches 712; Conservative 0; Mismatches 690; Indels 21; Gaps 6;

QY 32 aactcttcatcaacaatgaattcgtctcctccaagggtccgagagagattaaacgtcacga 91

Db 557 aacttttcatcaacaatgagttcgttgattcgtgtcgggcaagaccttggcacottca 616

QY 92 acccgtgggacgaatccaccgttggccactgatgttcacgtggccaacgcgcgatgtcg 151

Db 617 atcc--ggctacatccaaggagagattgtccaagtctccgaggggagataaggctgatatcg 673

QY 152 acagtgcaagtagccgcttcggtgcaggcgggtcaaaaaggccc--atggaagaagttca 208

Db 674 acctggctgtgaaaagcggccaagaagaagccttccatcgcgactcgggaatggcgcaaatga 733

QY 209 caggtgcacaaagcgcggcgttgcattgaagttcgcggaccctcgcgcgagaagaacgcgcg 268

Db 734 gtctctgcagcgcaccaaatttgatgaacaaacttgtgccctgatggatcgggacaagg 793

QY 269 agaagctcgcctcgttgagtcgctgccaccgcggtagaccggtgtcgatgatactcatt 328

Db 794 cattcctggccagcttggagaccaccagacaatggaaagccctacgtgag---gccctct 850

QY 329 tcgacattccaacacatggtctccggtgttgcgtactatgcaggtcgggcccgcacaagatcg 388

Db 851 tcgatgtgacctactcgatcctgacgtgcagtactacgcgcggtggaccgcacaagtctct 910

QY 389 ccggaaaagaccttcccaggagacaacgggcaacgcgaatggcgtttacgagccgatggggg 448

Db 911 tcggcgacaccattcccgcgtggagggttcttctccatgacccgcgaaggaaacccggtggcg 970

QY 449 tgtgtgctgtatgccagctggaaagcgactttcttcaagctggtgctggaagatagccc 508
Db 971 tgggtggccagatataccctggaaactatccctgctgctgctgctggaagtggggcc 1030
QY 509 cgcgcctgcgcgcgcgtgctcctctcatcttcaaaagcctggagaaatccccgtgggcg 568
Db 1031 ccgctctggcgtcgatgcaccatcatcatgaaccccgaggagcaaaactcctctgactg 1090
QY 569 tcttggcctgcctctctcttcgcgaagcgggattccctcctggagtcgtgagttcc 628
Db 1091 cccctcacatggctgttggcgaagaggcaggcttcccagcgggtgtgatcaatgtgg 1150
QY 629 tcaactgga--gcacgagtgacgggtggaagcattggcgtgcacatggacatggcgaaga 685
Db 1151 tcaacggatttgggtcccaactgcccgtgtgctgctatcagtgctcatcctgatattgccaagg 1210
QY 686 tcaactcacaaagatctgtcggcgggtggccgcgcgcgtcaagcaagcaacactcaagtcca 745
Db 1211 tggctttcacccggtccgtggaaatcggtcggtattgtcattgcaggtcgagccacgtcga 1270
QY 746 acatgaagcgcgtcaactctagaactgggggaaagcccaacacatcgtcttcaacgaagctc 805
Db 1271 acttgaagagggtctccctggaaatggcggcgaagagcccggttggcttctgagatgatg 1330
QY 806 ctctcgaaacggcagtcgggggaatcggaagagatttctcaaaaattcgggcaaatgtggg 865
Db 1331 cggacattgacttcgcgcgtggaaacacccacgaggctctgttctcc-----aatcacg 1384
QY 866 tccccccctcctgttctgtagtgcaatgggggaaatttagcggagaaatctccatggagtcc 925
Db 1385 gccagagctgtgtgcgggcagtcgcacctacgtttcacgaaaagatctacgacgagttcg 1444
QY 926 gtcatggctcatttggaggctgtcagagatggcttggccagaacccattggaaaccaaga 985
Db 1445 tggccaaaggcagctgccaaaggaaagcccgcaagggtgggcaatccttctcgagagaatg 1504
QY 986 ggacgcagtggtccctctcgtcgaaagttccagtagtacgacagagcttgggttaacattgacg 1045
Db 1505 tgcagcagggtcccccagatcgacgagtagacatgctgaccaaagggtgtgggtacatcgaga 1564
QY 1046 ttggc--aaggataccgcgcagctcctcactgctgcttggtagaaaggcgcaagggat 1102
Db 1565 gtggcaagaaggaggtgccaaagtgcaggtgtgtggcaagaggatcgccaatgtcgggt 1624
QY 1103 tgcgattgaacagcagatatttgcataatcccaacccagcgacgaaatgtgttgagg 1162
Db 1625 tottcgtgagcccaactgtcttctcgcagctgaagtagatgatacggtgcccaggagg 1684
QY 1163 agatctttggccccgtcttctccattgaagcgttcaagacggaagagagggccattgaga 1222
Db 1685 agatctttggccccgttccagtcgatcttcaagttcagctccctggaggagatgatcgtac 1744
QY 1223 ttgccaatgacacgacttatgggctagcctcgggtcatttataccaaatctctcaacaggg 1282
Db 1745 gggccaaacaatgtgcaaatatggcctggccgcggcgtttattaccaatgacatcaacaagg 1804
QY 1283 gtctccgtgtctgtcgcgtcgcagacccgggtggcgtctcgatcaacttcccctttatcc 1342
Db 1805 cactgaaattcgcacaacaatgtggatgcccggctccgtgtggatcaactgctatgatgccg 1864
QY 1343 ccgagacacaaactccgttttggcgggatgaacaaatcgggctcagggcagagagatgagcg 1402
Db 1865 tgctgcccctccaccccggttcgtgtggatacaagcactctggcattggcagggaattgggca 1924
QY 1403 aagaagggtcgaaggcgtacttggagcccaagaccattaatat 1445
Db 1925 aggatggcctggataactatcttggaaaccaagaccatcaccat 1967

RESULT 7
AAS70764
ID AAS70764 standard; cDNA; 3442 BP.

XX AAS70764;
AC 13-FEB-2002 (first entry)
XX DNA encoding novel human diagnostic protein #6568.
DE Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder; ss.
KW Homo sapiens.
OS WO200175067-A2.
XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US08631.
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
DR P-PSDB; ABG06577.
XX New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity -
PT Claim 1; SEQ ID No 6568; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 3442 BP; 920 A; 760 C; 864 G; 898 T; 0 other;
SQ Query Match 14.5%; Score 211.6; DB 23; Length 3442;
Best Local Similarity 50.5%; Pred. No. 6e-53;
Matches 678; Conservative 0; Mismatches 644; Indels 20; Gaps 6;
Qy 120 tgatgttcacgtggccaaacggccgagtgctgcagagtgacgttagccgttcggtgcaggc 179
Db 235 tgaagtggaaaggagagataagcccgacgtggacaaggctgtggaggctgcacaggtgc 294
Qy 180 gggtcaaaaaagggc---ccatggaaagagttcacaggtgcacaacgcgcggtgcatgct 236
Db 295 ctccagagggggtcgcgcacatggcggcggtggaggtccctgagtcgtggcggtgctgca 354
Qy 237 taagttcgcggaccctgcgcgagagaagacgcgcggaagctcgtcgtctggagtcgctgcc 296

Query Match		13.3%	Score 194.8;	DB 7;	Length 2700;
Best Local Similarity		52.2%;	Pred. No. 6.1e-48;		
Matches 639;		Conservative 0;	Mismatches 548;	Indels 36;	Gaps 8;
QY	37	ttcatcaacaatgaattcgtctcctccaagggtccgagagattaaacgctcacgaaccg	96		
Db	894	ttcatcaacaacgagttcgtgaaggcggttgaggccaagaccttcaggttcatacccc	953		
QY	97	tgggacgaatccaccggttgccactgatgttccactggtggccaaacggccgagatgcgacagt	156		
Db	954	tccaacga---gaaggtcatcacctccgtccacgaagccaccgagaaggtgttga--t	1007		
QY	157	gcagtagccgcttcgggtgcaggcggtcaaaaaaggcccatgggaagaagttcacagggtgca	216		
Db	1008	gtgcggtgctgctgcccgtgctgoccttggaggggccatggcgccaggtcacccccctct	1067		
QY	217	caacgcgcgcgctgcattgcttaagtctgcgcggacctcgcgcgagagaacgcgcgagaagctc	276		
Db	1068	gagcgtggcattttgatcaacaagctggcggtatctgtatggagcgtgatatacgacaccctc	1127		
QY	277	gctcgtctggagtcgctgcccacgcggtagacccggtgtcgtatgatacactcatttcgacatt	336		
Db	1128	gccgtatcagtcctctcgcacaacggcaaggctttcaccatggcca---aggtcgatctt	1184		
QY	337	ccaaaacatggtctccgtgtttcgtactatgcaggctggcgccgacaagatgcgcggaaaag	396		
Db	1185	gccaaactccattggttctgctgatactactacgtggtggcggaacaagattcacggtcag	1244		
QY	397	acctttcccaggagacaacggcaagc---cgaattggcgtttacgagccgcatgggggtgtgt	453		
Db	1245	accattgacacaacccccgagactcttacctacacccgcacgcgagcccggttggtgttgc	1304		
QY	454	gctggtattgccagctggaaacgcgactttcttacctgctggcgtggaagatagccccgcc	513		
Db	1305	ggtcagatcacccctggaaacttcccctctctgatgtgtggtcctggaagattggaccgcgt	1364		
QY	514	ctgcgcgcgcgctgctcctctcattcattcaaaagcctcggagaaaatcccgcgtggcgtctg	573		
Db	1365	gttgccgctggttaacactgtgttctcacaagaccgcccagcagacccctctctcgcgcctt	1424		
QY	574	ggcctcgtcctctctcttcgcagaaagccggattccctcctggtgagtcgtagtctcctcact	633		
Db	1425	tacgtgctaagctgatcaaggaggctccattccccgctggtgtgtagtacaacgtcatctct	1484		
QY	634	gg---agcacgagtgacgggtgaagcattggcgtgcacatggacattgcgaagatcagc	690		
Db	1485	ggctttggccgtaccgctggtgctgcctatctccagccacatggacatttgacaaggttgc	1544		
QY	691	ttcaacaagatctgtcggcgtggtggccgcgctcaagcaagcaacactcaagtccaacatg	750		
Db	1545	ttcactggctctactcttgttggacctaccatccatccgtcaggccgctgtctaagagcaacctg	1604		
QY	751	aagcgcgtcactctagaaactgggggaaaagcccaaccatcgtcttcaacgaagctcctctc	810		
Db	1605	aagaaggtcactctttagctcgtggtggcaagtctcccaacatcgtctttagatgatgctgac	1664		
QY	811	gaacggcagtcgggggaatcggcaaaaggat-----ttctcaaaattcgggcaaaattgg	864		
Db	1665	attgacaacgcacattctcctgggccaaactttggtatctcttcaaccacggccagtgctgc	1724		
QY	865	gtccccccctcctctgttgcgtagtgcaatggggaatttagcggagaaaattccatggagtc	924		
Db	1725	tgtgctggatcccgatcctcgtgtccaggagggtcatctacgacaagttcgtcgcgcgcttc	1784		
QY	925	cgtcatggctcatttggaggctgtcagagatggcttggccagaccccatgggaacccaag	984		
Db	1785	aaggagcgtgc-----ccagaagaacaaggctcggaaaccccttcgagcaggac	1832		
QY	985	aggacgcattggtccctcctcgtcagaaagtcocccagtcacgacagagtcttgggtaacattgac	1044		
Db	1833	accttcagggtcccccagggtttccacagctccagttccgacctatcatgtagtacaac	1892		
QY	1045	gttggc---aaggataccgcgcagctcctcactggcgttggtagaaaaggcgacaaggga	1101		

Db	1893	cacggcaagaaggctggtgctaccgtcgccaccggtggtgaccgccacggcaacgaggt	1952
QY	1102	ttcgcgattgaaccgacgacgatatttgtcaatcccaaacccaggcagcaaaatttggttgag	1161
Db	1953	tacttcacgcctactgtcttcacagacgtcacttccgacatgaagattgccaggag	2012
QY	1162	gagatctttggccccgtcttgtccattaaagacgttcaagacgggaagaagagccattgag	1221
Db	2013	gagatcttcggctcctgctgcactatccagaagtccaaggatgtggtgaggtatcaag	2072
QY	1222	attgccaatgacacgacttatgg	1244
Db	2073	atcgccaantcgaccgactacgg	2095
RESULT 9			
AAC78074			
ID	AAC78074 standard; cDNA; 3463 BP.		
XX	AAC78074;		
XX	08-FEB-2001 (first entry)		
XX	Human cancer associated gene sequence SEQ ID NO:468.		
DE	Human; cancer associated gene; cancer antigen; detection; cancer;		
KW	diagnosis; cytostatic; proliferative; vulnery; immunomodulator;		
KW	antidiabetic; antiasthmatic; antirheumatic; antiarthritic; antiviral;		
KW	antiinflammatory; antithyroid; antiallergic; antibacterial; cardiac;		
KW	dermatological; neuroprotective; thrombolytic; coagulant; neutropic;		
KW	vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;		
KW	immune disorder; haematopoietic cell disorder; autoimmune disorder;		
KW	allergic reaction; graft versus host disease; organ rejection;		
KW	haemostatic; thrombolytic; cardiovascular disorder; infection;		
KW	neurological disease; drug screening; ss.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200055350-A1.		
XX			
PD	21-SEP-2000.		
XX			
PF	08-MAR-2000; 2000WO-US05882.		
XX			
PR	12-MAR-1999; 99US-0124270.		
XX			
PA	(HUMA-) HUMAN GENOME SCI INC.		
XX			
PI	Rosen CA, Ruben SM;		
XX			
DR	WPI; 2000-587533/55.		
DR	P-PSDB; AAB43865.		
XX			
PT	Novel isolated nucleic acids comprising sequences encoding peptides		
PT	useful for treating or diagnosing e.g. cancer -		
XX			
PS	Claim 1; Page 1003-1004; 2352pp; English.		
XX			
CC	AAC77607 to AAC78448 encode the human cancer associated proteins given		
CC	in AAB43398 to AAB44239. The proteins can have activities based on the		
CC	tissues and cells the genes are expressed in. Example of activities		
CC	include: cytostatic; proliferative; vulnery; immunomodulator;		
CC	antidiabetic; antiasthmatic; antirheumatic; antiarthritic;		
CC	antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;		
CC	dermatological; neuroprotective; cardiac; thrombolytic; coagulant;		
CC	neutropic; vasotropic; antipsoriatic and antiangiogenic. The		
CC	polynucleotides and polypeptides can be used for preventing, treating or		
CC	ameliorating medical conditions and diagnosing pathological conditions.		
CC	Polynucleotides, polypeptides, antibodies, agonists and antagonists from		
CC	the present invention may be used to treat immune disorders by activating		
CC	or inhibiting the proliferation, differentiation or mobilisation of		
CC	immune cells, to treat disorders of haematopoietic cells, autoimmune		

XX Claim 1; Page 894-895; 2338pp; English.
CC AAF15566 to AAF16505 encode the human prostate cancer associated
CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
CC The prostate cancer antigens can have neuroprotective, cytostatic,
CC cardioactive, immunomodulatory, muscular, vulnery, gastrointestinal,
CC nephrotropic, antiinfective, gynaecological and antibacterial activities,
CC and can be used in gene therapy. The prostate cancer antigen
CC polynucleotides may be used for detection of prostate cancer, chromosome
CC identification, as chromosome markers, and for numerous other diagnostic
CC or research purposes. The prostate cancer antigens may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to
CC AAB57303 represent sequences used in the exemplification of the present
CC invention.
XX
SQ Sequence 1646 BP; 450 A; 397 C; 425 G; 371 T; 3 other;

Query Match 13.0%; Score 190; DB 21; Length 1646;
Best Local Similarity 52.7%; Pred. NO. 1.3e-46;
Matches 484; Conservative 1; Mismatches 422; Indels 12; Gaps 3;

QY 539 tcaagcctcggagaaatccccgctggcgcttctggccctcgctcctctcttcgcagaag 598
Db 15 tgaagtggcagagcagacccccctctctgcccctgtattngcctcctcatcaaggagg 74
QY 599 ccggaattccctcctggagtcgtgcagttcctcactgga---gcacgagtgacgggtgaag 655
Db 75 caggcttccccctggggtggaacatcatcacggggtatggcccaacagcaggtgcgg 134
QY 656 cattggcgctgcacatggacattgcgaagatcagcttcacaagatctgtcggcggtggcc 715
Db 135 ccacgcgccagcacatggatgttgacaaagtgccttcaccggttcacacgaggtggcc 194
QY 716 ggcgcgtcaagcaagcaacactcaagtcacatgaagcgcgtcactctagaaactggggg 775
Db 195 acctgatccagaaagcagctggcgattccaaacctcaagagagtcacccctggagctgggtg 254
QY 776 aaagcccaaccatcgtcttcaacgaagtcctcctcctcgaacggcagtcgggggaatcgcaa 835
Db 255 gtaagaccccgacatcgtcgtggccgattgctgacatggagcagtcgctggagcagtgcc 314
QY 836 aggatcttcaaaaattcgggcaaatgtgggtccccccctcctgtgttgctagtgcgaatggg 895
Db 315 acgaagccctgttcttc-----aacatgggcccagtgctgtgctgtggtctcccggacct 368
QY 896 gaaatrragcggagaaaattccatggagtcctcgttcatttgaggcgtgtcagagat 955
Db 369 tcgtggaagaatccatctacaatgagtttctcgagagaacccgtggagaaagcaagcaga 428
QY 956 ggcttgccagaaacccattggaacccaagaggacgcatggtcccttcgtcgcagaagtcctc 1015
Db 429 ggaagtggggaacccctttgagctggacacccacagcaggggcctcaggtggacaaggagc 488
QY 1016 agtaacagacagagctcttgggttaacattgacgtttggc--aaggataccgcgagctcctca 1072
Db 489 agttgaacgagtgctcaggtacatccacgcttggccagaaaggaggcgcaaaactcctct 548
QY 1073 ctggcgttggtagaaaaggcgagacagggtattcgcgattgaaccgacgatatttgtcaatc 1132
Db 549 gtcgaggagagcgtttccgggagcgtgtgttcttctcatcaagcctactgtcttbtggtggcg 608
QY 1133 ccaaacaggcagcaaaaatttggtttgaggagatctttggccccctgttctgcatttaaga 1192
Db 609 tgcaggatgacatgagaattgccaagaggagatctttggcctgtgtgcagccctgttca 668
QY 1193 cgttcaagacggaagaaaggccatttgagattgcgaatgacacgacttatgggctagcct 1252
Db 669 agttcaagaagattgagggaggtggttgagaggggccaacaacacacagggtatggcctggctg 728

QY 1253 cggtcatttataccaaaatctctcaacaggggtctcctcgtgtctcgtcggcgtcagagaccg 1312
Db 729 cggctgtgttcaccccggtatctggacaaggccatgtacttccccagcactccagggccg 788
QY 1313 gtggcgtctcgatcaacttcccttattatcccccgagacacaaaactccgttttgccggcatga 1372
Db 789 ggaccgtgtgggtaaacacacctacaacatcgtcacctgccaacacgcatatttgagggttta 848
QY 1373 aacaatcgggctcaggcagagagcagtcaggcgaagaaggcctcaaggcgtacttggagacca 1432
Db 849 aggaatctggaacgggagggtggtgggtgaggatgggcttaaggccttacacagaggttaa 908
QY 1433 agaccattaatatccacgt 1451
Db 909 agacgggtcacccatcaaggt 927

RESULT 11
AAZ33590
ID AAZ33590 standard; cDNA; 2036 BP.

XX AAZ33590;
XX 08-DEC-1999 (first entry)
XX Human breast tumour-associated EST 50.
DE Expressed sequence tag; EST; human; breast; cancer; cytostatic;
KW medicaments; gene therapy; treatment; fat metabolism; ss.
XX Homo sapiens.
OS DE19813835-A1.
PN 23-SEP-1999.
PD 20-MAR-1998; 98DE-1013835.
PR 20-MAR-1998; 98DE-1013835.
XX (META-) METAGEN GES GENOMFORSCHUNG MBH.

XX Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E, Rosenthal A;
XX WPI; 1999-528979/45.

XX Human nucleic acid sequences and protein products from normal breast
XX tissue, useful for breast cancer therapy -

XX Claim 3; 138; 206pp; German.

XX This invention describes novel human nucleic acid sequences from normal
XX breast tissue which have cytostatic activity. The nucleic acid sequences
XX can be used to produce and isolate full-length gene sequences. They can
XX be used to express proteins, which can be used as tools to find an
XX activity against breast cancer. The sequences can be used in sense or
XX antisense form. They are especially useful for medicaments for gene
XX therapy to treat breast cancer and for treating illnesses associated
XX with fat metabolism. AAZ33541-233610 represent expressed sequence tags
XX described in the method of the invention.

XX Sequence 2036 BP; 487 A; 529 C; 576 G; 444 T; 0 other;

Query Match 12.6%; Score 184.6; DB 20; Length 2036;
Best Local Similarity 48.4%; Pred. NO. 6.2e-45;
Matches 702; Conservative 0; Mismatches 729; Indels 20; Gaps 6;

QY 12 gcctgacgaatacaagagtgaaactcttcatcaacaagaatgcgtctcctccaagggtc 71
Db 158 gcccgaggtctctctgcaaccagattttcataaacaatgaatggcacgatgccgtcagcag 217
QY 72 cgagagattaaacgctcacgaacccgtgggacgaatccaccgttgccaactgatgttcacgt 131

Db 218 gaaacattccccaccgtcaatccgtccactcgagaggtcatctgtcaggtagctgaag 277
Qy 132 gccaacgcgcccagatgtcgacagtgacagtcgcgcttcggtgcagggcgtcaaaaaggg 191
Db 278 ggaaca--ggaagatgtggacaaggcagtgaaagccgcccggcgccctccagctggg 334
Qy 192 ccaatggaagaagtccaaggtgcacaacgcgcggtgcgtgcattctaa---gttcgcgga 248
Db 335 ctcaacttggccgcgtggacgcgtacacacaggggcccggctgctgaaccgctggccga 394
Qy 249 cctgcgcgagaagaacgcgagaagctcgcgtcgtctgtgagtcgctgcccacccggtagacc 308
Db 395 tctgatcgagcgggacccgacctacctggcgcccttgagacccctggacaactggcaagcc 454
Qy 309 ggtgtcgatgatcactcatttcgacattccaacatggtctccgtgttctcgctactatgc 368
Db 455 ctatgtcatctctacctggtggtattggacatggtctcctcaaatgtctccggtattatgc 514
Qy 369 aggtgggcccgaacaagatgcgggaaagaccttccc--gaggacaacggcaagccgaa 425
Db 515 cggtcgggctgataagtaccacgcggaaacacatccccattgacggagacttcttcagcta 574
Qy 426 ttggcgttacgagccgcatgggggtgtgtgctggtattgccagctggagcgcgacttttct 485
Db 575 cacagcccatgaacctgtgggggtgtgcgggcagatcattccgtggaatttcccgctcct 634
Qy 486 ttaegtgcgtggaagatagccccccgcgcgcgcgtcctcctccttcattcttcaaaagc 545
Db 635 gatcaagcatggaagctgggcccagcctctatgtggccaacctgatacgaaggctggctt 694
Qy 546 ctggagaaaatccccgcgtggcggttctgtgggcccgcgcgcgtcctccttcagaaagc 605
Db 695 agctgagcagacacccctcacccgcctctatgtggccaacctgatacgaaggctggctt 754
Qy 606 cctcctggagtcgtgcagttcctcactggagcacgagtgacgggtgaa--gcattggc 662
Db 755 tccccctggtgtggtcaacattgtgcctggatttggccccacgcgtgggcccgcctatgc 814
Qy 663 gtgcacatggacattgcaagatcagcttcacaagatctgtcggcggtggccgcgcgt 722
Db 815 ctcccatgaggtgtggacaagtggcattccacaggctccactgagattggccgcgtaat 874
Qy 723 caagcaagcaacactcaagtccaacatgaagcgcgtcactctagaaactgggggaaagcc 782
Db 875 ccaggttgctgctgggagcagcaacctcaagagagtgaccttgagctggggggaagag 934
Qy 783 aacatcgtcttcaacgaagctcctctctcgaaacggcagtcgggggaatcgggcaagattt 842
Db 935 cccaacatcatcatgtcagatgccgatattggaattggccgtggaacagggcccacttcgc 994
Qy 843 ctcaaaaattcgggcaaatgttggtccccccctcctgtttgctagtgaatggggaattt 902
Db 995 cctgttcttcaacca-----gggccagtgtgtgtgtgcgggtcccggaaccttcgtgca 1048
Qy 903 agcggagaaaattccatggagtcctcgtcatggctcatttggaggctgtcagagatggcttg 962
Db 1049 ggaggacatctatgatgagtttgttgagcgggaggggttgcggggcccaagtctcgtgtgt 1108
Qy 963 ccagaaaccttggaaaccaagaggacgcgcgtgtcccttcgtcgacaagtcctccagtagca 1022
Db 1109 cgggaacccctttagtagcaagaccgagcagcagggcgccgaagtggatgaaactcagtttaa 1168
Qy 1023 cagagtcttgggtaacattgacgttggcaaggtataccgcg--cagctcctcactggcgtt 1080
Db 1169 gaagatcctcggctacatacaacacgggggaagcaagagggggcgaagtgtgtgtgtggg 1228
Qy 1081 ggtagaaaggcgcacaagggatttcgcgattgaaacgcagatatattgtcaatcccaacca 1140
Db 1229 ggcattgtctgtgacccgttgttacttctcagccactgtgtttggagatgtgcaggat 1288
Qy 1141 ggcagcaaaaatttggtttgaggagatcttggccccgcgtcttgtccattaaagacgttcaag 1200

Db 1289 ggcatgaccatcgccaaggaggagatcttcggggccagtgatgcagatcctgaagttcaag 1348
Qy 1201 acggaagaagagggccattgagattgccaatgacacgacttatgggctagcctcggtcatt 1260
Db 1349 accatagaggaggttgttgggagagccaacaattccacgtacgggtggccgcagctgtc 1408
Qy 1261 tataccaaatctctcaacaggggtctcctggtctcgtcgtcgcgctcgagaccggtggcgtc 1320
Db 1409 ttcaaaaaggatttggacaaggccaattacctgttcccaggccctccaggcggcactgtg 1468
Qy 1321 tcgatcaacttcccccttattcccccgagacacaaaactccgttttggcgcatgaaacaatcg 1380
Db 1469 tgggtcaactgctatgtatgtgttggagcccagtcaccttgggtgtacaaagatgtcg 1528
Qy 1381 ggctcaggcagagagtaggcgaagaagggtcacaagcgtacttggagccccaaagaccatt 1440
Db 1529 gggagtgccgggagttggcgagtagtcgggtcgaggcatcacactgaagtgaaaactgtc 1588
Qy 1441 aatatccacgt 1451
Db 1589 acagtcaaaagt 1599

RESULT 12

AAF82082

ID AAF82082 standard; DNA; 1541 BP.

AC AAF82082;

DT 26-JUN-2001 (first entry)

DE Human aldehyde dehydrogenase ALDH2 nucleotide sequence SEQ ID NO:3.

KW Aldehyde dehydrogenase; glycerol dehydratase; 3-HP; glycerol;
feedstock; 3-hydroxypropionic acid; genetic engineering; glucose;
bacterial host; absorbable prosthetic device; surgical suture;
beta-lactam; acrylic acid; trifluoromethylated alcohol; diol;
polyhydroxyalkonate; copolymer; lactic acid; ds.

OS Homo sapiens.

FH Key Location/Qualifiers

CDS 22..1521

FT /*tag= a

FT /partial

FT /product= "aldehyde dehydrogenase ALDH2"

XX WO200116346-A1.

XX 08-MAR-2001.

XX 30-AUG-2000; 2000WO-US23878.

XX 30-AUG-1999; 99US-0151440.

XX (WISC) WISCONSIN ALUMNI RES FOUND.

XX Suthers PF, Cameron DC;

XX WPI; 2001-315988/33.

XX P-PSDB; AAB74924.

PT 3-Hydroxypropionic acid preparation, for use e.g. as monomer, by
fermenting recombinant microorganisms expressing genes for suitable
enzymes in the presence of glycerol or glucose -

PS Disclosure; Page 29-32; 63pp; English.

XX The present invention describes a method for the production of

CC 3-hydroxypropionic acid (3-HP). The method comprises fermenting a
recombinant microorganism in the presence of a source of glycerol (I)

CC or glucose, where the microorganism: (i) expresses genes for non-native
enzymes which catalyse the production of (3-HP) from (I); (ii) carries

QY 1190 agacgttcaagacggaagaagaggccattgagattgccaatgacacgacttatgggctag 1249
 |||
Db 1416 tgaagttcagtgatgtggatgaggtgataaagaggcgacgagacgaagtacgggctag 1475
 |||
QY 1250 cctcggtcatttatataccaaatctctcaacagggtctctcgtgtctcgtcggcgtcgaga 1309
 |||
Db 1476 cggcaggggttttcacgaagaatctggacacggaacagggttttcaagggtttgaaag 1535
 |||
QY 1310 ccggtggcgtctcgatcaacttccctttatccccgagacacaaaactccgttttggcgca 1369
 |||
Db 1536 ctggtaccgtatgggttaactgtctcgacgtatttgatgcagccataccatttgggtgtt 1595
 |||
QY 1370 tgaacaatcgggctcaggcagagagctaggcgaagaagggtctcaaggcgtaacttgagc 1429
 |||
Db 1596 acaagatgagtggaatgggagagagaaaggcatatacagttctcaataattacttgaga 1655
 |||
QY 1430 ccaagac 1436
 |||
Db 1656 tcaaggc 1662

Search completed: April 27, 2002, 00:40:59
Job time: 18893 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2002, 23:02:13 ; Search time 7811.87 Seconds
(without alignments)
3921.782 Million cell updates/sec

Title: US-09-882-694A-4
Perfect score: 1464
Sequence: 1 atggtttcttcgcctgacga.....tccacgtcaacatagatga 1464

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_ba:*
- 2: gb_htg:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vi:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_sts:*
- 28: em_un:*
- 29: em_vi:*
- 30: em_htg_hum:*
- 31: em_htg_inv:*
- 32: em_htg_other:*
- 33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description

1	1464	100.0	1464	6	AX076845
2	270.4	18.5	1653	8	UMU74468
3	267.6	18.3	1488	6	A43196
4	267.6	18.3	1660	8	CHCLAH3
5	258.8	17.7	1659	8	AF275347
6	240.6	16.4	1751	8	AB037421
7	240.4	16.4	3021	9	BC001619
8	240.4	16.4	166207	9	AL135785
9	239.2	16.3	1485	6	A43275
10	239.2	16.3	1647	8	AAALTA2
11	238.8	16.3	6074	9	HUMMTALD
12	238.2	16.3	79554	8	NCB11B23
13	232	15.8	2236	10	BC020001
14	223.2	15.2	2060	10	MMU96401
15	219.2	15.0	1563	10	AF280404
16	217.8	14.9	2048	3	EBALDH
17	217.4	14.8	2024	10	RATALDH
18	216	14.8	2372	10	AF246711
19	214.4	14.6	3263	10	AF152359
20	214.4	14.6	3390	10	AF253409
21	211.6	14.5	3442	9	HSU07919
22	209.8	14.3	1770	10	AF001896
23	209.8	14.3	2035	10	AF001897
24	209.8	14.3	2095	10	AF001898
25	209.6	14.3	1722	10	MMRALDH2
26	209.6	14.3	1855	8	AB030939
27	208.2	14.2	2012	10	RNU79118
28	206.8	14.1	9179	1	PSP296087
29	206.6	14.1	1806	10	RATALDHA
30	203.2	13.9	2244	5	AF339837
31	202.2	13.8	2745	8	AF260123
32	202	13.8	11197	1	AE009326
33	202	13.8	11339	1	AE008288
34	201.6	13.8	2233	5	AF315691
35	201.6	13.8	2240	10	RNU60063
36	201.4	13.8	1630	9	AB015226
37	200.6	13.7	2745	8	AF260124
38	200.6	13.7	2745	8	AF260125
39	200.6	13.7	2745	8	AF260126
40	198	13.5	3217	8	PAU40996
41	196.8	13.4	1889	10	RNADR
42	194.2	13.3	2804	8	EMEALDA
43	193.2	13.2	1393	9	AB015227
44	192.8	13.2	3092	9	AB015228
45	191.4	13.1	1868	8	AF269064

ALIGNMENTS

RESULT	1				
AX076845	AX076845	Sequence	4	from Patent	WO0105980.
LOCUS	AX076845	Sequence	1464 bp	DNA	linear
DEFINITION	AX076845	Sequence	4	from Patent	WO0105980.
ACCESSION	AX076845	Sequence	1464 bp	DNA	linear
VERSION	AX076845.1	GI:13121519			
KEYWORDS	AX076845.1	GI:13121519			
SOURCE	AX076845.1	GI:13121519			
ORGANISM	AX076845.1	GI:13121519			
REFERENCE	AX076845.1	GI:13121519			
AUTHORS	AX076845.1	GI:13121519			
TITLE	AX076845.1	GI:13121519			
JOURNAL	AX076845.1	GI:13121519			
FEATURES	AX076845.1	GI:13121519			
source	AX076845.1	GI:13121519			
BASE COUNT	AX076845.1	GI:13121519			

TITLE	Direct Submission	
JOURNAL	Submitted (05-JUN-2000) CAB, University of Maryland, 5115 Plant Sciences Bldg., College Park, MD 20742, USA	
FEATURES	Location/Qualifiers	
source	1..1659	
	/organism="Cladosporium fulvum"	
	/strain="Cf4"	
	/specific_host="Lycopersicon esculentum Mill."	
	/db_xref="taxon:5499"	
gene	1..1659	
	/gene="aldh"	
CDS	61..1554	
	/gene="aldh"	
	/note="pSI-9; ALDH"	
	/codon_start=1	
	/product="aldehyde dehydrogenase"	
	/protein_id="AAF82789.1"	
	/db_xref="GI:9082283"	
	/translation="MSLSVQLTPTSAKYEQPTGLFINNEWKPVSKGFDDVINPTDES SVICVAEADKDVDAIAAARAKAFEGPWRQETPENRGKLLVKLAELFEKNLDLFASV KALDNKAFNMAKVDIGMCAGCLRYGGWADKIEGKVDTTPDTFNYIKKEPIGVCVG IIPWNFLLMWAKIGPAVATGNTVIKTAEQTPLGAYVAANLKEAGPPGVVNVVIS GFGVAGAAIAASHMDIDKVAFTGSTVGTITLKAAGSNLKKVTLELGKSPNIVFDD ADIDNAISWNVFGIFFNHGQCCGASRIYVQEGIYDKFLQRFXRAQKNVVGDPFAAD TFQGPQSVQVDFRIMGYIQACRDAGATVEIGGNRKGDKGIFYEPTIFANVTEDMKIM QEETFGPVCISISKFTKEDAIKVGNAATPYGLAAAVHTRNLNLAIEVSNALRAGTVVWN TYNTLHQLPFGGFKESGIGRELGEDALANYTQTKTVSIRLGDALEFG"	
BASE COUNT	381 a	496 c 440 g 342 t
ORIGIN		

BASE COUNT	381 a	496 c	440 g	342 t	
ORIGIN	i i n l d h h o l p f g g f k e s g i g r e g e d a l a n y t o t k t v s i r l g d a l f g "				
Query Match	17.7%; Score 258.8; DB 8; Length 1659;				
Best Local Similarity	52.8%; Pred. No. 7.8e-53;				
Matches 763; Conservative	0; Mismatches 647; Indels 36; Gaps 8;				
QY	17	acgaatacaagagtgaaactcttcatcaacaataatcgctctcctccaaaggggtccgaga	76		
Db	104	ACGAGCAGCCCACTGGCCTTTTCATCAACAATAGTGGTGAAGCCAGTCAGCGGCAAGT	163		
QY	77	gattaaagctacgaacccgtgggacgaatccacacgttgccactgatgttcacgtggcca	136		
Db	164	TCTTCGATGTCATCAACCCAACCGACGAGT--CCGTCTCTGCCAAGTCGCAGAGGCCG	220		
QY	137	acgcgccgatgtcgcagtgtagcgcgttcgggtgcaggcgggtcaaaaagggcccat	196		
Db	221	ATGAGAAAGGATGTGACATCGCGTCGCGCGCGCCGCAAGGCCTTCGA--GGGCCCAT	277		
QY	197	ggaagaagttcacagggtgcacaaacgcggcggtgcatgcttaagttcgcggacctgcgcg	256		
Db	278	GGAGACAAGAGACCCAGAGAACCGCGGCAAGCTCCTCGTGAAGCTCGCCGAGCTCTTCG	337		
QY	257	agaagaacgcgcgagaaagctcgctcgctggagtcgctgcccacccggtagaccggtgcga	316		
Db	338	AGAAGACCTCGACCTCTTCGCCCTCCGTCAAGGCGCTCGACAACG---GAAAGGCGTTCA	394		
QY	317	tgatcaactcatttcgacattccaaacatggctcccggtgttcgctactatgcaggctggg	376		
Db	395	ACATGGCCAAAGTTGACATTGGCATGTGCGCTGGCTGCCTCCGATACTACGGTGGCTGGG	454		
QY	377	ccgacaagatcgccggaagacatttcccgaaggacaaacggcgaagccgaattggcgt---	433		
Db	455	CCGACAAGATTGAGGGCAAGGTGCTCGACACCACCTCCAGATACCTTCAACTACATCAAGA	514		
QY	434	acgagccgatgggggtgtgtgtcgtggtattgccagctggaacgcgacttttcttaactcg	493		
Db	515	AGGAGCCAATCGGTGTGTGCGGTGAGATCATCCCATGGAACCTCCCACTCCTCATGTGGG	574		
QY	494	gctggaagatagccccccgcttcgcccgcggctgctccttcatcttcaaaagcctcgaga	553		
Db	575	CATGGAAGATTGGTCCAGCCGTTGCCACTGGTAACACCGGTGTTCATCAAGACTGCTGAGC	634		
QY	554	aatccccgcgtgggcgttctgggcctcgctcctctcttcgcagaaagccggattccccctcg	613		

/note="CpG island"
/evidence=not_experimental
join(12875..13334,23096..23205,23946..24062,25750..25899)
/gene="bAl13024.1"
join(12875..13334,23096..23205,23946..24062,25750..25899)
/gene="bAl13024.1"
/note="match: CDNAS: Em:AB006141
match: proteins: Tr:O88812 Tr:Q16270"
/codon_start=1
/evidence=not_experimental
/product="bAl13024.1 (similar to insulin-like growth
factor binding protein)"
/protein_id="CAD13245.1"
/db_xref="GI:17384405"
/translation="MPRLSLLPLLLLLLLPLPPLSPSLGIRDVCGRRPKGQPCRP
GCPAPAPCPAPGISALDECGCCARCGLGAGASCGRAGRCGPGLVLCASQAAGAAPEG
TGLCVCAQRGTVCSDGRSYPSCALRLRARTPRAHPGHLHKARDGPCEPAPVVVP
PRSVHNTGAOVLSCSEVRAPVPTVTRKVKTSPEGTQALEELPDGHDVNIADVVRGG
PSDHEATAWILINFLRKEDEGVYQCHAAANMVGAEESHSTVTVLDLSKYSRSHFPAPDD
RM"
repeat_region 12890..12940
/note="17 copies 3 mer ctg 86% conserved"
repeat_region 13062..13139
/note="26 copies 3 mer gcg 66% conserved"
repeat_region 13692..13787
/note="4 copies 24 mer 85% conserved"
repeat_region 14013..14091
/note="MER69 repeat: matches 1..78 of consensus"
repeat_region 15994..16286
/note="AluJo repeat: matches 3..303 of consensus"
repeat_region 16338..16403
/note="MIR repeat: matches 76..146 of consensus"
repeat_region 16444..16747
/note="AluY repeat: matches 1..300 of consensus"
repeat_region 17142..17406
/note="AluJo repeat: matches 34..298 of consensus"
repeat_region 17407..17457
/note="17 copies 3 mer gga 92% conserved"
repeat_region 18914..18997
/note="MIR repeat: matches 112..183 of consensus"
repeat_region 19236..19756
/note="MER68A repeat: matches 1..561 of consensus"
repeat_region 19842..19967
/note="MIR repeat: matches 34..188 of consensus"
repeat_region 20015..20199
/note="MIR repeat: matches 7..241 of consensus"
repeat_region 20328..20622
/note="AluJo repeat: matches 10..299 of consensus"
repeat_region 20641..20780
/note="L1ME1 repeat: matches 6017..6150 of consensus"
repeat_region 22762..22896
/note="MIR repeat: matches 5..140 of consensus"
repeat_region 23235..23264
/note="15 copies 2 mer gt 100% conserved"
repeat_region 24244..24733
/note="MLT1D repeat: matches 8..505 of consensus"
repeat_region 26799..27098
/note="AluY repeat: matches 1..299 of consensus"
misc_feature complement(27151..27534)
/note="match: GSS: Em:AQ131150"
misc_feature complement(27159..27539)
/note="match: GSS: Em:AQ128717"
misc_feature 27545..28093
/note="match: GSS: Em:AQ763725"
misc_feature 27546..28104
/note="match: GSS: Em:AQ768584"
repeat_region 28476..28558
/note="MIR repeat: matches 65..143 of consensus"
repeat_region 28848..29167
/note="AluJo repeat: matches 1..302 of consensus"
repeat_region 29434..29574
/note="MIR repeat: matches 87..242 of consensus"
repeat_region 29929..30009

/note="MIR repeat: matches 77..155 of consensus"
complement(30372..30771)
/note="match: STS: Em:G23339"
complement(31296..31713)
/note="match: GSS: Em:AQ683842"
31311..31411
/note="L2 repeat: matches 2319..2418 of consensus"
complement(31523..31902)
/note="match: GSS: Em:AQ130619"
33086..33279
/note="MLT1I repeat: matches 63..265 of consensus"
33439..33742
Query Match 16.4% Score 240.4 DB 9 Length 166207
Best Local Similarity 49.8% Pred. No. 3.2e-48
Matches 725; Conservative 0; Mismatches 716; Indels 15; Gaps 4
QY 5 ttcttgcctgacgaataacaagagtgaactcttcatacaaatgaattcgtctcctcca 64
Db 41465 TTCTGAACCCAGACATCCCTACAACACGCTGTTTCATCAACAATGAATGGCAAGATGCAG 41406
QY 65 aggggtccgagagattaaagctcacgaacccggtgggacgaatccaccgttgccactgatg 124
Db 41405 TCAGCAAGAAGACCTTCCCGACGGTCAACCCCTACCACCGGGAGGTCAATGGGCACGTGG 41346
QY 125 ttcacgtggccaaacgcgccgatgtcgacagtgacagtagcagcttcggtgcagggcggtca 184
Db 41345 CTGAAGGTGACCGGGCTGATGTGGATCGGGCCGTGAAGGACGCGCGGAAGCCTTCCGCC 41286
QY 185 aaaagggcccatggaagaagttcacaggtgcacaaacgcggtgagtgcatgcttaagtctg 244
Db 41285 TGGGTCCCATGGCGCGGATGGATGCTCTCTGAGCGGGCGGCTGCTGAACCGCCTGG 41226
QY 245 cggacctcgcgagagaagacgcgagagaagctcgctcgctctggtgagtcgctgcccaccggt 304
Db 41225 CAGACCTAGTGGAGCGGGATCGAGTCTACTTGGCTCTACTCGAGACCTTGGACAATGGGA 41166
QY 305 gaccggtgctgatgatcactcatttcgacattccaaacatggttcctcgtgttcctact 364
Db 41165 ACCCTTTCCAAGAGTCTTACGCCTTGGACTTGGATGAGGTGATCAAGGTGTATCGGTACT 41106
QY 365 atgcaggctgggcccagacagatcgccggaagacaccttcacgagagacacacggaagcga 424
Db 41105 TTGCTGGCTGGGTGACAAAGTGGCATGGCAAGACCATCCCATGATGGCCAGCATTTCT 41046
QY 425 att--ggcgttacgagccgatgggggtgtgtgtgctggttatggccagctggaacgcgactt 481
Db 41045 GCTTCAACCGGCATGAGCCCGTGGTGTCTGTGGCCAGATCATCCCGTGAACCTTCCCT 40986
QY 482 ttctttacgtcggctggaagatagcccccgccctcgcccggttcctccttcatcttca 541
Db 40985 TGGTCATGCAGGGTTGAAACTTGCCCCGGCACTCGCCACAGGCAACACTGTGGTTATGA 40926
QY 542 aagcctcggagaaatccccgctggggtgtcttgggacctcctcctcttcgcgagaagcgg 601
Db 40925 AGTGGCAGAGCAGACCCCTCTCTGCCCTGTATTTGSCCTCCCTCATCAAGGAGSCAG 40866
QY 602 gattccctcctggagtcgtgcagttcctcactgg---agcacgagtgcggttggaagcat 658
Db 40865 GCTTCCCTCGGGTGGTGAACATCATCACGGGGTATGCCCCAACACAGCAGGTGCGGCCA 40806
QY 659 tggcgtcgcacatggacattgcgaagatcagcttcacaagatctgtcggcggtgcccgcg 718
Db 40805 TCGCCAGCACGTGGATGTGACAAAGTTGCCCTTCACCGGTTCCACCGAGGTGGGCCACC 40746
QY 719 ccgtcaagcaagcaacactcaagtccaacatgaagcgcgtcactcttagaactgggggaaa 778
Db 40745 TGATCCAGAAAGCAGCTGGCGATTCCAACCTCAAGAGAGTCAACCTGGAGCTGGGTGTA 40686
QY 779 agccaaccatcgtcttcaacgaagctcctctctcgaacggcagtcgggggaatcggcgaagg 838
Db 40685 AGAGCCCCAGCATCGTGTGCTGCGCGATGCTGACATGGAGCATGCGCGTGGAGCAGTCCACG 40626

Db 996 GACAC-----TCCAGGGTCTCAAGTCTCGCAGCTCCAGTTCGACCGT 1038

QY 1027 gtcttgggttaacattgacgttggaagga---taccgcgcagctcctcactgcgcttgggt 1083

Db 1039 ATCATGGGCTACATCGAGGAGGGCAAGAGTCTGGCGGACCATCGAGACTGGTGGCAAC 1098

QY 1084 agaaaggcgacaaggattcgcgattgaaccgacgatatattgtcaatcccaaacaggc 1143

Db 1099 CGTAAGGGTGACAAGGGTTACTTTCATCGAGGCCACAAATCTTCCAAAGTAAACCGAGGAC 1158

QY 1144 agcaaaatttggttgaggagatctttggcccccgtcttgtccattagacgttcaagacg 1203

Db 1159 ATGAAGATTTCAGCAAGAAGAGATCTTCGGGCCCGTCTGCACAATCTCCAAGTTCAAGACA 1218

QY 1204 gaagaagaggccattgagattgccaatgacacgacttatgggctagcctcggtcatttat 1263

Db 1219 AAGGCCGACGTCATCAAGATTGGCAACAACACACACATACGGTCTTTCGGCCCGTGTACAC 1278

QY 1264 accaaatctcacaaggggtctccgtgtctcgtcgcgctogagacgggtggcgctcgc 1323

Db 1279 ACATCCAACTCACCACTGCCATCGAAGTTGCCAACGCGCTCCGTGCAGGAACGTGCTGG 1338

QY 1324 atcaacttcccctttatccccgagacacaaactccgtttggcgcatgaacaacatcgggc 1383

Db 1339 GTCAACTCTFACAACACTCTTCACTGGCAGCTTCCCTTCGGAGGGTACAAGGAGTCTGGT 1398

QY 1384 tcaggcagagagctaggcgaagaaggctcaaggcgctacttggagcccaagaccattaat 1443

Db 1399 ATTGGCGCGAGTGGGAGAGGGCGCGCTGGACAACATACATCCAGACCAAGACCGTGTCT 1458

QY 1444 atcc 1447

Db 1459 ATTC 1462

RESULT 10

AAALTA2

LOCUS AALTA2 1647 bp mRNA linear PLN 27-FEB-1997

DEFINITION A.alternata Alta2 mRNA for aldehyde dehydrogenase.

ACCESSION X78227

VERSION X78227.1 GI:467614

KEYWORDS aldehyde dehydrogenase.

SOURCE Alternaria alternata.

ORGANISM Alternaria alternata

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Dothideomycetes; Pleosporales; Pleosporaceae; mitosporic Pleosporaceae; Alternaria.

REFERENCE 1 (bases 1 to 1647)

AUTHORS Achatz,G., Oberkofler,H., Lechenauer,E., Simon,B., Unger,A., Kandler,D., Ebner,C., Prillinger,H., Kraft,D. and Breitenbach,M.

TITLE Molecular cloning of major and minor allergens of Alternaria alternata and Cladosporium herbarum

JOURNAL Mol. Immunol. 32 (3), 213-227 (1995)

MEDLINE 95206305

REFERENCE 2 (bases 1 to 1647)

AUTHORS Achatz,G.

TITLE Direct Submission

JOURNAL Submitted (18-MAR-1994) G. Achatz, Dept of Genetics and General Biology, Hellbrunnerstrasse 34, 5020 Salzburg, AUSTRIA

FEATURES

source

1. 1647

/organism="Alternaria alternata"

/strain="08-0203-Berlin"

/db_xref="taxon:5599"

/clone_lib="lambda Zap II"

/dev_stage="hyphea, conidiospores"

59. 1546

/EC_number="1.2.1.3"

/note="Alta2"

/codon_start=1

/product="aldehyde dehydrogenase (NAD+)"

/protein_id="CAA55071.1"

/db_xref="GI:467615"

/db_xref="SWISS-PROT:P42041"

CDS

/translation="MTSVKLSTPQTGEFEQPTGLFINNEFKAVDGTKTFDVPINPSTEE
VICSVQEATEKDVDIAVAARKAFNGPWAKETPENRKLNLKLDLFEKNADLIAAVE
ALDNGKAFSMAKNVDVPAAGCLRYGGWADKIEGVVDTAPDSFNIRKSLLVFAVR
SSMELPILMWSWKIGPAIATNTVVLKTAEQTPLSAYIAKLIQEAGFPFGVINVITG
FGKIAGAAMSAMDIDKIAFTGSTVVGRIKMSAAGSNLKKVLTLELGGKSPNIVFADA
DLDEAIHWVNEGIYFNHGQACAGSRIYVQEEIYDKFIQFKERAQNAVGDPPFAATL
QGPQVSQLQDFRIMGYIEGKKSGATIEETGNRKGDKGYFIEPTIFSNVTEDMKIQQE
EIFGPVCTISKEFTKADVIKIGNNTTYGLSAAVHTSNLTTAIEVANALRAGTVWVNSY
NTLHWQLPFGGYKESGIGRELGEAALDNYIQTKTVSIRLGDVLFQ"

BASE COUNT 407 a 450 c 417 g 373 t

ORIGIN

Query Match 16.3%; Score 239.2; DB 8; Length 1647;
Best Local Similarity 51.5%; Pred. No. 5e-48;
Matches 744; Conservative 0; Mismatches 663; Indels 37; Gaps 7;

QY 18 cgaatacaagagtgaactcttcaacaataatgaattcgtctcctccaaaggggtccgagag 77

Db 100 CGAGCAGCCCAACCGGACTCTTCATCAACAATAGATTTCGTAAAGGCTGTTGACGGCAAGAC 159

QY 78 attaaogctcaagaaacccgtgggacgaatccaccgttgccactgatgttcacgtggccaa 137

Db 160 CTTTGATGTTATCAACCCCTCCACTGAGGAGGTTCATCTGCA---GTGTTACGAGGCCAC 216

QY 138 cgcggccgatgtcgacagtgacagtagccgcttcgggtgcaggcgggtcaaaaagggcccatg 197

Db 217 CGAGAAGGATGTTGACATTGCTGTTGCTGCCGCCCGCAAGCCTTCAAC---GGCCCCATG 273

QY 198 gaagaagttcacaggtgcacaacgcgcggcggtgcattgcttaagttcgcggacccctgcgcga 257

Db 274 GGCAAAGGAGACACACAGAGAAGAGGGGAAAGCTGCTTAACAAGCTTGCCCGATCTGTTCTGA 333

QY 258 gaagaacgcgcgagaagtcgctcgtctggtgagtgctgctgccaccggttagaccggtgtcgat 317

Db 334 GAAGAATGCCGACCTCATTTGCTGCTGCTGAGGCTCTCGACAACGGCAAGGCTTCAGCAT 393

QY 318 gatcactcatttcgacattccaaacatggtctcgcgtgttctcgtctactactgcaggctgggc 377

Db 394 GGCCAAGAACGTCGATGTTCCCGCCCGCTGCTGCTGAGGTACTACGGAGGATGGGC 453

QY 378 cgacaagatgcgcggaagacaccttcccgcgaggacaaacgcgaagccgaattggcgttacga 437

Db 454 CGACAAGATTGAGGGCAAGGTCTGTCGACACAGACACCCGACAGCTTCAACTACATCCGCA 513

QY 438 gccgatgggggtgtgtggtattgccagctggaacgcgacttttcttactcgtcggtg 497

Db 514 GAGCCTATTGTTGTTGCGGTTCAGATCATCCATCGAACTTCCTATTCTCATGTGGTTCATG 573

QY 498 gaagatagccccgcctcgcgcgcggtgctcctcctcatcttcaaaagccctcgagaaaac 557

Db 574 GAAGATTGGTCTCTGCCATCGCCACTGGTAACACCGTGTGCCATGTCCGCTCACATGGA 633

QY 558 cccgctggggttctgggcctcgcctcctcctctcttcgcagaagccggattccctcctggagt 617

Db 634 ACCTCTCTCCGCATACATTGCCTGCAAGCTGATCCAGGAGGCCGCTTCCACACAGGTGT 693

QY 618 cgtgcagttcctcactggagcacgagtga---cgggtgaagcattggcgtgcacatgga 674

Db 694 CATCAACGTCATCACTGGTTTCGGAAAGATCGCCGGTGTGCCATGTCCGCTCACATGGA 753

QY 675 cattgcgaagatcagcttcacaagatctgtcggcggtggcgcgccgctcaagcaagcaac 734

Db 754 CATTGACAAGATTGCCTTTACTGGTTCAACCGTGTTCGGCCGTCAAAATCATGAAGTCTGC 813

QY 735 actcaagttccaacatgaagcgcgtcactcttagaactgggggaaagcccaacca-----t 788

Db 814 CGCTGGCTCCCACTTGAAGAAGGTCACTCTTGAGCTCGGAGGCAAGAGCCCCAACATTGT 873

QY 789 cgtcttcaacgaagctcctctcgaacgcgcgagtcggggggaatcggcaaggaatttctcaaa 848

Db 874 CTTGCCCGACCGCAGATCTTGACGAGGCTATCCACTGGGTCAACTTTGGTATTACTTCAA 933

Db 3071 TGATCCAGAAAGCAGCTGGCGGATTCCAACCTCAAGAGACTCACCTGGAGCTGGGTGGTA 3130

QY 779 agccaaccatcgtcttcaaacgaagctcctctctcgtcgaacggcagtcggyggaatcggcgaagg 838

Db 3131 AGAGCCCCAGCATCGTCTGGCCGATGCTGACATGAGCATGCCGTGGAGCAGTGCACAG 3190

QY 839 atttctcaaaaattcgggcaaaatttgggtccccccctcctctgttctgtagtgcaatgggga 898

Db 3191 AAGCCCTGTTCTTC-----AACATGGGCCAGTGTGCTGTGCTGGCTCCCGGACCTTCG 3244

QY 899 atttagcggagaaattccatggagtcctcgtcatggtcatttggagctgtcagagatggc 958

Db 3245 TGAAGAATCCATCTACAAATGAGTTTCTCGAGAGAAACCGTGGAGAAAGCAAAGCAGAGGA 3304

QY 959 ttggccagaacccattggaacccaagaggacgcatggtcccttcgtcgcacaagtcccaagt 1018

Db 3305 AAGTGGGGAACCCCTTTGAGCTGGACACCCAGCAGGGGCCCTCAGGTGGACAAAGCAGCAGT 3364

QY 1019 acgacagagtccttgggtaacattgacgtttgc---aaggataccgcgcagctcctcactg 1075

Db 3365 TTGAACGAGTCTTAGGCTACATCCAGCTTGGCCAGAAAGAGGGCGGCAAAACTCTCTGTG 3424

QY 1076 gcgttggttagaaaggcgacaaagggattcgcgattgaaccgacgatatttgcataatccca 1135

Db 3425 GCGGAGAGCGTTTCGGGGAGCGTGGTTTCTTCATCAAGCCTACTGTCTTTGGTGGCGTGC 3484

QY 1136 aaccaggcagcaaaaatttggtttgaggagatcttbtggcccgctcttgcatttaagacgt 1195

Db 3485 AGGATGACATGAGAAATTGCCAAAGAGGAGATCTTTGGGCCTGTGCAGCCCTGTTCAAGT 3544

QY 1196 tcaagacggaagaagagccattgagattgccaatgacacgacttatgggctagcctcgg 1255

Db 3545 TCAAGAAGATTGAGGAGGTGGTTGAGAGGGCCAAACACCCAGGTATGGCCTGGCTGCGG 3604

QY 1256 tcatttatacacaatctctcaacaggggtctccgtgtctcgtcgccgctcagacccggtg 1315

Db 3605 CTGTGTTCAACCCGGGATCTGGACAAGGCCATGTACTTCACCCAGGCACCTCCAGGCCGGA 3664

QY 1316 gcgtctcgatcaacttccccctttatccccgagacacaaactccgttttggcgcatgaaac 1375

Db 3665 CCGTGTGGGTAAACACCTACAACATCGTCACTGCCACACGCCCATTTGGAGGGTTTAAAG 3724

QY 1376 aatcgggctcaggcagagagctaggcgaagaagggtcactcaaggcgtacttggagcccaaga 1435

Db 3725 AATCTGGAACCGGGAGGAGCTGGGTGAGGATGGGCTTAAGGCCCTACACAGAGGTTAAAGA 3784

QY 1436 ccattaatatccacgt 1451

Db 3785 CGGTCAACCATCAAGGT 3800

RESULT 12

NCB11B23/C

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Neurospora crassa.

Neurospora crassa

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

Sordariales; Sordariaceae; Neurospora.

1 (sites)

Schulte,U., Algn,V., Hoheisel,J., Brandt,P., Fartmann,B.,

Holland,R., Nyakatura,G., Mewes,H.W. and Mannhaupt,G.

Unpublished

2 (bases 1 to 79554)

German Neurospora genome,project.

Direct Submission

Submitted (18-JAN-2002) MIPS, Institut fuer Bioinformatik,

GSF-Forschungszentrum fuer Umwelt und Gesundheit, GmbH,

Ingolstaedter Landstrasse 1, D-85764 Neuherberg, FRG, E-mail:

G.Mannhaupt@gsf.de Project Coordinator: Ulrich Schulte, Institute

COMMENT

of Biochemistry, Heinrich-Heine-University, D-40225 Duesseldorf,
E-mail: ulrich-schulte@uni-duesseldorf.de
BAC clone 11B23(strain OR74A) is available at the Fungal Genetic
Stock Center, http://www.fgsc.net
Sequencing was performed by MWG Biotech AG, Ebersberg, Germany,
http://www.mwgdna.com
Information on performance of analysis and a more detailed
annotation of this entry and other sequences of linkage groups II
and V can be viewed at: http://mips.gsf.de/proj/neurospora.

FEATURES

source

1..79554

/organism="Neurospora crassa"

/db_xref="taxon:5141"

/chromosome="6"

misc_feature

1..21781

/note="overlap to BAC B19C19"

1078..1653

/gene="B11B23.010"

complement(join(1078..1231,1478..1653))

/gene="B11B23.010"

complement(join(1078..1231,1478..1653))

/gene="B11B23.010"

/note="similarity to Elongin C transcription elongation
factor (YPL046c), Saccharomyces cerevisiae, PIR:JC5792"

/codon_start=1

/product="related to Elongin C transcription elongation
factor"

/protein_id="CAD21119.1"

/db_xref="GI:18376341"

/translation="MDATNGNGGASKYITLVSKDGFVYVLRREATLCSDIYIRGMLRNN
MTEARTGRCELODINGVILEKVVEYFHYWYKYRDRDVPDMEIPVENCLELVAAADYL
GMSDKPH"

exon

complement(1078..1231)

/gene="B11B23.010"

/number=1

complement(1232..1477)

/gene="B11B23.010"

/number=1

complement(1478..1653)

/gene="B11B23.010"

/number=2

join(4122..4226,4295..4498,4563..5009)

/gene="B11B23.030"

/note="strong similarity to RHO1 PROTEIN, Candida
albicans, SWISSPROT:RHO1_CANAL
Contains ATP/GTP-binding site motif A (P-loop) [GDGGCGKT]
contains EST gb:BG279563, BG279562"

/codon_start=1

/product="probable RHO1 PROTEIN"

/protein_id="CAD21120.1"

/db_xref="GI:18376342"

/translation="MTEGPAYSKVVVVVGDCGCKTCLLISYSGYFPEKYVPTVFEN
YITYPTHPTGKTVELALWDTAGQEYDRLRPLSYPETDLIFVCFAIDCPNSLENVMD
KWYPEVLHFCPYTPLILVGLKSDLRNKKTCDMLKTQGLTPVTTEQGLAVAKKMGACQ
MECSSKEMKGVEEIFEQAILTVAANDRKTLDEQEAANGMLGVGAGSGSGKSGISFSSG
DKAGSGIGPVKAAAGVGTIVPKTRKKRKCMM"

4122..4226

/gene="B11B23.030"

/number=1

4122..5009

/gene="B11B23.030"

4227..4294

/gene="B11B23.030"

/number=1

4295..4498

/gene="B11B23.030"

/number=2

4499..4562

/gene="B11B23.030"

/number=2

4563..5009

/gene="B11B23.030"

/number=3

exon

4122..4226

/gene="B11B23.030"

/number=1

4122..5009

/gene="B11B23.030"

4227..4294

/gene="B11B23.030"

/number=1

4295..4498

/gene="B11B23.030"

/number=2

4499..4562

/gene="B11B23.030"

/number=2

4563..5009

/gene="B11B23.030"

/number=3

```
gene      5791. .7688
/ gene="B11B23.040"
exon      complement(5791. .6147)
/ gene="B11B23.040"
/ number=1
gene      complement(join(5791. .6147,6331. .6937,7024. .7688))
/ gene="B11B23.040"
CDS       complement(join(5791. .6147,6331. .6937,7024. .7688))
/ gene="B11B23.040"
/ note="similarity to chromosome 5, p1 clone:mx10,
arabidopsis thaliana, TREMBL:ATAB5248_2"
/ codon_start=1
/ product="conserved hypothetical protein"
/ protein_id="CAD21121.1"
/ db_xref="GI:18376343"
/ translation="MSFLNLNLNLNTYQNHFRHPSSSSPPPIPLPPGISRSFI
PTPHGDLEILHTATTTSSSQQPALFFIHGGMGSASVWLEYLSFFSEQQQQQNI
PCYAISMRGHGGWPSYLRMVYGTTKGLTGDVVGIRWAVDRERGQRKLOQLGRE
GIKQGGKEGRGEWEVEMVLIGHSSGGGLAQYILSAGLLGLGVKIKGLVLMGAVP
GFGSIPYLTWFRDPLFALRLLHLHGNPSPLSHPSLVRLIFSSNDLLPDSYLSRF
MSFMSRYESLLWPLGMARSFADPKVLLRISGWEETTTKQEQKQGEERLVIAGORD
KIMTQPI MORLADWYRAAYQORLVVEKKIDGMVMSGENETVDEEVEVEVEGSGGF
KDTAGRGVRFALVPGAGHCCNDVGVGARKPVCLLNTLRKWCAGRWVAILRDEYG
VPMRSRVGETSHKVIQGVIPPVKVHLITDSLLOLTHTGISSHATSMPTINILLSFVAS
AISFDCLSTALSAGIKPNTDTERITYTKPTGTQK"
intron    complement(6148. .6330)
/ gene="B11B23.040"
/ number=1
exon      complement(6331. .6937)
/ gene="B11B23.040"
intron    / number=2
/ complement(6938. .7023)
/ gene="B11B23.040"
/ number=2
exon      complement(7024. .7688)
/ gene="B11B23.040"
/ number=3
exon      8798. .8857
/ gene="B11B23.050"
/ number=1
gene      8798. .10714
/ gene="B11B23.050"
CDS       join(8798. .8857,8921. .10714)
/ gene="B11B23.050"
/ note="similarity to lariat-debranching enzyme, YKL1149c,
Saccharomyces cerevisiae, PIR:A38491"
/ codon_start=1
/ product="related to lariat-debranching enzyme"
/ protein_id="CAD21122.1"
/ db_xref="GI:18376344"
/ translation="MSIDIQTTTQGVRIAVEGCGHGTNLAIYAATEKSAKERGWGV
DLIIIGDFQAVRNAADLNAMSVPKYRELGDHEIYSGVRKAPYLTFVAGNHEAAS
HLWELYGGWVCNIIYMGAAVRLGLRLAGMSGIWKGYNYRAHHERLPFNDDDDV
KSPYHVREIDVRKLLQQTQVDIGISHDWPRAIKHHGNTKHLWNMKPDFERESADGSL
GNQAAEYVMDRLRPPYWFSAHLCKYSAIKTYDLPKETTSDGGETTKPEETAVPQOE
QOPAPAAANPDEIDLMEDEPAPAPQQAAPVATSNPDEIDLDDDEVAPTAPSQPAPT
SAAPAAVTEEDPTEALRSLLPASESNPNQSSYNAOPHGCNSGNKPKPGQVPPTI
TNTTVRFLSLDKCLPKGRKFLQALAEIEPINSPKELQRLRLSYDPDWLSITLAFHPLL
RSVAGKHCPSPHPNAPIVDLGEAGYKPLIEARKWVDEHITGDKLLVPEDFVQTAPPI
DLVNRPETGNMTDEMPEVTNPHTAEFCRLLGVENYWDASEEEREGRVQGP RPMAER
FGGRGGHRRGGGGGGGGWRGGRGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG"
intron    8858. .8920
/ gene="B11B23.050"
/ number=1
exon      8921. .10714
/ gene="B11B23.050"
/ number=2
rRNA      10924. .11043
/ product="5S ribosomal RNA"
gene      11719. .12591
/ gene="B11B23.060"
exon      complement(11719. .12591)
/ gene="B11B23.060"
```

```
gene      / number=1
complement(11719. .12591)
/ gene="B11B23.060"
CDS       complement(11719. .12591)
/ gene="B11B23.060"
/ codon_start=1
/ product="hypothetical protein"
/ protein_id="CAD21123.1"
/ db_xref="GI:18376345"
/ translation="MATPTPTTNNKDGTDPTSLARTIAHMKDHATDLLLLHMLIHYS
PLALLQLPLDLKPVRLVDLRGEDVDLRVDMSLAELKVOTRAKEGAPOTRTTIAH
TIPIHPPMKSWAERRERLVAMTKEARGLGLPVEGVASEEGKKKEQEKVGKREKGG
GGAVIDTYLAPLRFWDGAVAISVAFFFGCYAAVKGGYVEQGLMAWEFTQKWWPVGGAD
GFKWLVNAMFWPVLGIHIAEMVMFERTLAKNDKILMGGLTWLKMVSVFFEGYMAFR
RYDEKAAELERKQH"
gene      complement(13073. .14179)
/ gene="B11B23.070"
13073. .14179
gene      / gene="B11B23.070"
exon      complement(13073. .14179)
/ gene="B11B23.070"
/ number=1
CDS       complement(13073. .14179)
/ gene="B11B23.070"
/ codon_start=1
/ product="hypothetical protein"
/ protein_id="CAD21124.1"
/ db_xref="GI:18376346"
/ translation="MVSAAEPSFRSESNKFLYSHGPFKIQRTHQIPIVMALPVST
VALEIFVDANPEVWFSRLRDOAHNCNDLNRAKAAVRHREIFLVDLLQKHVGKPDGHYLI
SAQTHPSFPVRLARAMYKDKSAIWCLPHCSDDITLSKEEKEVDLTKRCRAQVETMSEY
DMVRMHHEEGWIEEAGRAPVTPGPEWQVSLREEVDEAIVSAMIVETKICORFQEVRR
RAKISPGTLPLWEELEKTEEEIRKVKLLGLGLSLNLSNRNRSQPLKVLMDPIR
AYLSATVHLFWATANLFLQASSSCLEIPKPNKAKVRSDDGGILRRYVAAVLCAEKGSQ"
Query Match 16.3%; Score 238.2; DB 8; Length 79554;
Best Local Similarity 53.3%; Pred. No. 1.1e-47;
Matches 694; Conservative 0; Mismatches 583; Indels 26; Gaps 8;
QY 37 ttcatacaacaatgaattcgtctcctccaaaggggtccgagagagattaaagctcacgaacccg 96
||||| ||||| || ||||| | ||||| ||||| ||||| ||||| |||||
Db 25397 TTCATAAACAAACGAGTTCGTTCGAGGGTGTGCACAAAGAGACTTCGAGGTCATCAACCCC 25338
QY 97 tgggacgaatccaccgttgccactgatgttcacgtggccaaagcggccgatgtcgacagt 156
||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 25337 GCCACCGAGGAGGTTCATCTGCTC---CGTCCACGAGGCCACCGAGAGGATGTCGATATC 25281
QY 157 gcagtagccgcttcggtgcaggcgggtcaaaaagggcccatggaagaagttcacagggtgca 216
||| ||||| || ||||| ||||| ||||| ||||| ||||| |||||
Db 25280 GCTGTGCGCGCTGCCCGCAAGGCTTTC---GAGGGTGTCTGGCCGACGTCACCTCCCCAG 25224
QY 217 caacgcgcggcgtgcattgcttaagtccggacacctcgcgcgaagaagaacccgagaagctc 276
||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 25223 CAGCGTGGCATCTACCTCCTCAAGCTCGCCGACCTGCTCGAGAAAGAACTCGACCTCCTT 25164
QY 277 gctcgtctggagtcgctgccaccggttagaccgggtgtcgtgatgatcactcatttcgacatt 336
||| ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 25163 GCCGCTGTGAGTCTCTCGACAATGGCAAGTCCATCACCATG---GCCCGTGGCGATGTG 25107
QY 337 ccaaacatggtctccgtgtttcgtactatgcaggctgggcgacaagaatcgcgcggaag 396
||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 25106 GGCGCCGCTCGTGGCACTATCCGCTACTACGGCGGTTGGCCGACAAAGATCGAGGGCAAG 25047
QY 397 acctt---tcccgaggacaacggcgaagcgaattggcgttacagcgcgatgggggtgtgt 453
||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 25046 ACCATTGATATCTCCCCCGACTCTCTCCACTACACCGCAAGAGCCTCTCGGTGTGTGC 24987
QY 454 gctggattgccagctggaacgcgacctttcttacctcgtcgggtggaagatagcccccgcc 513
||| || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 24986 GGTCAAGATCATCCCTGGAACCTCCCTCCTCATGCTTGCCTGGAAGGTCGSCCCCGCC 24927
QY 514 ctgcgcgcggcgtgctctctcatcttcaaaagcctcggagaaaatccccgcgttcttg 573
||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Db 24926 CTGGCCACCGGTAACACTATCGTCTATGAGACCGGCTGAACAGACGGCCTCTCTGCTCTC 24867

QY 574 ggcctcgctcctctcttcgcagaagccggattccctcctcgagtcgagtcagttcctcact 633

Db 24866 GTCTTTGCTCAGTTCGTCAAGGAGCGCGGTTCCCTCCTGGCGTTCTCAACATCATCTCC 24807

QY 634 gg--agcacgagtgacgggtgaagcattggcgctgcacatggacattgcgaagatcagc 690

Db 24806 GGTTCGGCAGAAATTGCTGGTCCGCCCATGGCCTCCACATGGACATGACAAGGTTGCC 24747

QY 691 ttcaacaagatcgtcgcgcggtggcgccgctcaagcaagcaaacactcaagtccaacatg 750

Db 24746 TTCACCGGCTCTACCATGCTCGCGCGCCCAATCATGAAGGCTGCCCGCGAGTCCAACTG 24687

QY 751 aagcgcgtcactctagaactggggggaagcaaacaccatcgtcttcaacgaagctcctctc 810

Db 24686 AAGAAGGTTACCTCGAGCTTGGTGGCAGTCCCAACATTATCTCAACGACGCCGAT 24627

QY 811 gaacggcagtcgggggaatcggcaaaaggatttctcaaaattcgggcaaaatttggtcccc 870

Db 24626 ATCGACCAAGGCCATTGATGGGTCAACTTCGGTATCTACTTCAACGACGGCCAGACTGC 24567

QY 871 ccctcctgtttgtagtgcaatgggggaatttagcgggagaaattccatggagtcgctcat 930

Db 24566 TGGCTGCTTCGTGTATACG-----TCCAGGAGGGTATCTATGACAAGTTTGTGGC 24514

QY 931 ggctcatttggaggtgtcagagatggcttggccag-aacccattggaacccaagaggac 989

Db 24513 GGCCTTCAAGCAGCGCGCCAGCAGACAAAGGTGGCGGACCCCTTCCACGACGAGACCTT 24454

QY 990 gcatggctccttcgtcgacaagtcctcagtaacagagatcttgggttaacattgacgttgg 1049

Db 24453 CCAGGGCCCCAGGTGAGCCAGCTCCAGTACGACCGCATCATGGTTACATCAAGGCCGG 24394

QY 1050 ca---aggataccgcgcagctcctcactcgttggtagaaaggcgacaaggattcgc 1106

Db 24393 TAAGGAGGAGGGTGCTACCGTGGAGACTGGTGGTGAGCGCCACGGTGACAAGGGCTACTT 24334

QY 1107 gattgaaccgagatatttgcataatcccaaacaggcagcaaaaatttggttgaggagat 1166

Db 24333 CATCCAGCCCAACCATTTTCACCAACGTCGGTCACGACATGAAGATCATGAAGGAGGAAT 24274

QY 1167 ctttggcccgctcttgcattaaagacgttcaagaacggaaggaagagggccattgagattgc 1226

Db 24273 TTTCCGCCCCCGTCTGCGCCGTGCGCAAGTCTCCACCGAGGAGGAGTTATCAAGCTCGG 24214

QY 1227 caatgacacgacttatgggtagcctcgtcatttatataccaaatctctcaacaggggtct 1286

Db 24213 CAACGACTCCCACTACGGCCCTTGCTGCTGCCGTCCACACCAAGGACCTCAACACTGCCAT 24154

QY 1287 ccgtgtctcgtcggcgctcgagacccggtggcgctcgtcgtatcaac 1329

Db 24153 CCGCGTCAGCAACCACTTAGGGCTGGTACCGTCTGGGTCAAC 24111

RESULT 13

BC020001

LOCUS BC020001 2236 bp mRNA linear ROD 22-JAN-2002

DEFINITION Mus musculus, aldehyde dehydrogenase 1 family, member B1, clone MGC:28561 IMAGE:4207479, mRNA, complete cds.

ACCESSION BC020001

VERSION BC020001.1 GI:18043460

KEYWORDS MGC.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 2236)

AUTHORS Strausberg,R.

TITLE Direct Submission

JOURNAL Submitted (19-DEC-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USDA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgapbs-re@mail.nih.gov

Tissue Procurement: Jeffrey E. Green, M.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center

Center code: BCM-HGSC

Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>

Contact: amg@bcm.tmc.edu

Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M., Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M., Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAK Plate: 37 Row: g Column: 23

This clone was selected for full length sequencing because it passed the following selection criteria: Similarity but not identity to protein.

Identity Location/Qualifiers

1..2236

/organism="Mus musculus"

/db_xref="LocusID:72535"

/db_xref="taxon:10090"

/clone="MGC:28561 IMAGE:4207479"

/tissue_type="Colon, normal. 5 month old male mouse."

/clone_lib="NCI CGAP Co24"

/lab_host="DH10B"

/note="vector: pCMV-SPORT6"

62..1621

/codon_start=1

/product="aldehyde dehydrogenase 1 family, member B1"

/protein_id="AAH20001.1"

/db_xref="GI:18043461"

/translation="MLTARLLPRLCLLOGRTTSYSTAAALPNPPIPCYNKLFIN NEWHDVSKKTFPTVNPVTGEVIGHVAEGDRADVLDLAVKAAREAFRLGSPWRMDASE RGRLLNLADLVERDRVYLASLETLDNGKPFQESYVLDLDEVIKVRYFAGWADKWHIG KTIPTMDGEHFCFTRHEPVGCGQIIPWNFPLVMQGWKLAPALATGNTVVMKVAQTPL SALYLASLIKEAGFPDPVWNIITGYPTAGAAIAQHMVDVKVAFSTGEVGHUQKAA GESNLKRVTLLEGGKSPSIVLADADMEHAVDQCHEALFFNMQCCGCCSRTFVEESII REFLERTVEKAKQRKVNPFELDTQQGPQVDKEQFERILGYIRLQKEGAKLLCGGER LGERGFFIKPTVFGVDQGMRIAKEEIFGPVQPLEFKKIEEVIQIRANTRYGLAANV FTRDLDAIYFTALQAGTWVNTYNTIVTCTPFGGFKESGNRELGEDGLRAYTEVK TVTIKVPEKNS"

BASE COUNT 562 a 571 c 644 g 459 t

ORIGIN

Query Match 15.8%; Score 232; DB 10; Length 2236;

Best Local Similarity 51.0%; Pred. No. 3e-46;

Matches 684; Conservative 0; Mismatches 640; Indels 18; Gaps 5;

QY 122 atgttcacgtggccaacgcgccgagtcgacagtgacagtcagtcggtcagcg 181

Db 267 ATGTGGCCGAAGGTGACCGGGCAGATGTGGATCTGGCTGTAAAGCAGCCGAGAAGCCT 326

QY 182 tcaaaaaagg--cccatggaagaagttcacaggtgcacaacgcgcgctgcatgctta 238

Db 327 TCCGCCCTGGGTTCCCATGGCGCAGATGGATGCCCTCAGAGCGGGCGCGCTGCTGAACC 386

QY 239 agttcggcgacctcgccgagaagaacgcccgaagctcgtctcgtgagtcgctgccc 298

Db 387 GCCTAGCTGATCTTGTGGAACGAGATCGAGTGTACTTGGCCTCACTGGAGACGCTAGATA 446

QY 299 ccggtagaccggtgtcgtatgatcactcatttcgacattcccaacatggtctccgtgttc 358

Db 447 ACGGGAACCTTTCCAGGAGTCTTATGTCTTGGATCTGGATGAAGTCATCAAGGTGTACC 506

QY 359 gctactatgcaggtcgccgacaagatcgccggaagaccttcccaggagacaacggca 418

Db 507 GTTACTTCGGCTGGCTGGCTGACAAAGTGGCATGGTAAGACCATCCCTATGGATGGTGAGC 566

QY 419 agccgaatlgg---cgttacgagccgcatgggggtgtgtgtgtattgccagctggaacg 475

Db 567 ATTTCTGCTTACCCGACATGAGCCAGTGGTGTCTGTGCCAGATAATCCCTTGGAACT 626

QY 476 cgacttttttacgtcggtggaagatagcccccgccctcgcccgcggtgctccttea 535

Db 627 TCCCACTGGTTCATGCAGGGCTGGAAGCTGGCCCCCGGCACTGCCACGGGCACACTGTGG 686

QY 536 ttttcaaagcctcggagaaaaatccccgclgggcgtttctgggcccctcgctcctcttcttcag 595

Db 687 TCATGAAGTGGCAGAGCAGACCCCACTCTCTGCTCTGTACTTGGCCTCCCTCATCAAG 746

QY 596 aagccggtattccctcctgagtcggtgcagttcctcactgg---agcaagagtgaacgggtg 652

Db 747 AGCGGGGTTTCCCCCAGGAGTGGTGAACATCATCACTTGGCTACGGCCCCACGGGGGAG 806

QY 653 aagcattgctgcacatggacattgcgaagatcgaaagatcagcttcacaagatctgtcggcgtg 712

Db 807 CTGCCATCGCTCAGCACATGGATGTGGATAAAGTCGGCTTTCACGGGCTCCACTGAGGTAG 866

QY 713 gccgcgcgtcaagcaagcaaacactcaagtcacaatgaaagcgcgtcactctagaactgg 772

Db 867 GCCACCTGATTCAGAAAGCAGCTGGCGAGTCTAACCTCAAGAGAGTCAACCCCTGGAGCTGG 926

QY 773 gggaaaaagcaaacatcgtcttcaacgaagctcctctgaaagcgcagtcggggaatcgg 832

Db 927 GTGGGAAGAGCCCCAGCATTTGTGCTGGCAGACGCTGACATGGAGCATGCCGTAGATCAGT 986

QY 833 caaaggatttctcaaaaattcgggcaaaattgggtccccccctcctgtttgctagtgaat 892

Db 987 GTCACGAAGCCCTTTCTTC-----AACATGGGCCAGTGTCTGTGCAGGCTCCCGGA 1040

QY 893 ggggaaatttagcgagagaaattccatggagtcctcgatcctcatttggaggctgcaga 952

Db 1041 CATTCTGTGAAGAGTCCATCTACCGTGAATTTCTCGAGAGAACTGTGGAGAAAGCCCAAGC 1100

QY 953 gatggcttgccagaaacccattggaaccccaagaggagcagtcggtcctctcgtcgacaaat 1012

Db 1101 AGAGGAAGTGGGGAAACCCCTTTGAGTTGGACACCCAGCAGGAGACCTCAGGTGGACAAGG 1160

QY 1013 ccagttacacagagctcttgggttaacattgaactgttg---caaggataccgcagctcc 1069

Db 1161 AGCAGTTTGAACGAATCCTGGGCTACATCCGGCTGGGACAGAAAGGAGGCGCAAGCTTC 1220

QY 1070 tcactggcgttggttagaaagggcgacaagggattcgcatggaaccgacgatatttgtca 1129

Db 1221 TCTGTGGCGGGAGCGTTTGGGGAGCGCGGCTTCTTCATCAAAACCCACAGTCTCGGGG 1280

QY 1130 atcccaaacaccaggcagcaaaaatttggttgaggagatctttgycctcgtctgtccatta 1189

Db 1281 ACGTTTCAGGATGGCATGAGGATCGCCCAAGGAGGAGATCTTTGGCCCGTGCAGCCTCTGT 1340

QY 1190 agacgttcaagacggaagaagagggccattgagattgccaatgacacgacttatgggctag 1249

Db 1341 TCAAGTTCAAGAAGATCGAGGAAGTAATCCAGAGAGGCCAACAAACACCAGGTATGGCCTGG 1400

QY 1250 cctcgggtcatttataccaaaatctctcaacaggggtctcgtgtcgtcgcgctcgaga 1309

Db 1401 CTGGCGGCTGTCTTACCCGAGACCTGGACAAGGCCATCTACTTCACGCAGGCCCTGCAAG 1460

QY 1310 ccggtggtcgtctcgatcaacttcccttttatcccccgagacacaaaactcgttttggcgga 1369

Db 1461 CTGGGACGGTGTGGTGAACACCTATAACATTGTACCTGCCACACGCCATTTCGGAGGCT 1520

QY 1370 tgaacaacatcggggtcagggcagagagctagcggaagaggggtcctcaaggcgtacttggagc 1429

Db 1521 TTAAGGAATCTGGCAATGGCAGGGAGCTGGGGAGGAGGACGGGCTCAGAGCCCTACACGGAGG 1580

QY 1430 ccaagaccattaatatccacgt 1451

Db 1581 TGAAGACTGTACCATCAAGGT 1602

RESULT 14

MMU96401

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

CDS

MMU96401

Mus musculus aldehyde dehydrogenase Ahd-2-like mRNA

2060 bp

linear

ROD 31-JUL-1997

U96401

U96401.1

GI:2289239

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 2060)

Hsu,L.C., Chang,W.C. and Yoshida,A.

Direct Submission

Submitted (03-APR-1997) Biochem. Genetics, Beckman Res. Inst. of

the City of Hope, 1450 E. Duarte Road, Duarte, CA 91010, USA

Location/Qualifiers

1..2060

/organism="Mus musculus"

/strain="C57BL"

/db_xref="taxon:10090"

/tissue_type="liver"

45..1550

/codon_start=1

/product="aldehyde dehydrogenase Ahd-2-like"

/protein_id="AAB64411.1"

/db_xref="GI:2289240"

/translation="MSSPAQPAVPAPLANLKHIOHTKIFINNEHDSVSSKKFPVLNPA

TEEVICHVEEGKADVDKAVKAAQAFQIGSPWRTMDASERHLNKLADLMERDRLL

LATMESNMAGKVFAHAYLLDVEISIKALOYFAGWADKIHGQTIPSDGNIFTYTRREPI

GVCQIIPWNGPLIIFTWKLGPALSCGNTVVVKPAEQTPALHMASLIKEAGFPFGV

VNIVPGYPTAGGAISSHMDIDKVSFGSTEVGKLKEAAKSNLKRVTLELGGKSPC

IVFADADLDSAVEFAHQGVFFHQGQICVAASRLFVEESIYDFVRSVERAKYILGN

PLNSGINQGPQIDKEQHNKILGLIESKKEGAKLECGGGRWGNKGFVQPTVFSNVTD

EMRIAKEEIFGPVQIIMKFKMDDVIKRANNTTYGLAAGVFTKDLDKAITVSSALOAG

MVVVNCYLAVPVQCPFGGFKMSGNGRELGEHGLYEYTELKTVMQISQKNS"

580 a 443 c 511 g 526 t

BASE COUNT

ORIGIN

Query Match 15.2%; Score 223.2; DB 10; Length 2060;

Best Local Similarity 49.8%; Pred. No. 4.3e-44;

Matches 709; Conservative 0; Mismatches 698; Indels 17; Gaps 5;

QY 35 tottcatcaacaatgaattcgtctcctccaaggggtccgagagagattaaacgctcacgaacc 94

Db 112 TCTTTATAAACAATGAATGGCATGATTCGGTGAGCAGACAAGAAATTTCTGTCTTAACC 171

QY 95 cgtgggacgaatccaccgttgccactgatgttcacgtggccaacgcggccgatgtcgaca 154

Db 172 CTGCAACTGAGGAGGTCATCTGCCACGTGGAAGAGGGGACAAAGGCTGATGTTGACAAAG 231

QY 155 gtgcagtagccgcttcggtgcagggcggtcaaaaagggcccatgggaagaagttcacagggtg 214

Db 232 CTGTGAAGCGCGCAAGACAGGCTTTCAGATTGGCTCCCATGGCGCACCATGGATGCTT 291

QY 215 cacaacgcgcgcgtgcgtcatgcttaagttcgcggacctcgccgagagaagaacgcgcgagaagc 274

Db 292 CAGAGAGGGCGCCCTGCTGAACAAGCTGGCTGACTTAATGGAGAGAGATCGTCTGCTGC 351

QY 275 tcgctcgtctggagtcgctgcccaccggtagaccgggtgtagaccgggtgtagatcactcatttcgaca 334

Db 352 TGGCTACAATGGAATCGATGAATGCTGGGAAAGTCTTTCCTCATGCATACCTGTGGATG 411

QY 335 ttcccaaacatggtctcctcgtgttttcgctactatgcaggctgggcccgcagacaagatcgccggaa 394

Db 412 TAGAGATCAGCATAAAGCATTACAGTACTTTCGACAGGCTGGGCTGACAAGATCCATGGCC 471

QY 395 agacctttccc---gaggacaacggcgaagccgaattggcgttacgagccgatgggggtgt 451

Db 472 AAACAATACCCAGTGTGTAACATATTCACCTTATACAAGGCGGTGAACCTATATGGGGTGT 531

QY 357 tcgctactatgcaggctgggcccgcacaagatcgccggaaagacacctttccc---gaggacaa 413
Db 440 CAGATATTTGCGGGTGGGCAGACAAAATCCAGGCGAGGACCATTCCCACAGATGACAA 499
QY 414 cggcaagccgaattgcggttacgagccgatgggggtgtgtgctggtattgccagctgaa 473
Db 500 CGTTGTGTGCTTCACCAGGCATGAGCCCATCGGGGTGTGTGGGGCCATTACACCATGGAA 559
QY 474 cgcgacttttctttacgtcggctggaagatagcccccgcctcgcgcggtgctcctt 533
Db 560 CTTCCCTGCTGATGCTGGCTTGGAACTGGCTCCTGCTCCTGCTGTGTGGGAACACCGT 619
QY 534 catcttcaaaagccctcgagaaaatccccgcgtggcggttctgggacctcgctcctcttcgc 593
Db 620 GGTCTGAAGCCAGCTGAGCAGACCCCTCTCACGGCTCTGTACCTAGCCTCTCTCATCAA 679
QY 594 agaagccggattccctcctgggagtcgtgcagttcctcactggagcacgagtgc---ggg 650
Db 680 AGAGGTCGGGTTCCCTCCGGGTGTGGTGAACATTGTACAGGCTTTGGGGCCACTGTGGG 739
QY 651 tgaagcattggcgtcgcaatggacattgcgaagatcagcttcacaagatctgtcggcgg 710
Db 740 AGCAGCAATTTCTCTCCCATCCGAGATCAACAAGATAGCCTTCACCGGCTCCACAGAGGT 799
QY 711 tggccgcgcgtcaagcaagcaactcaactcaagtccaaactgaagcgcgtcactctagaact 770
Db 800 TGGAAAGCTGGTCAGAGAGAGCGCCCTCCCGGAGCAACCTGAAGAGGCTCACACTGGAGCT 859
QY 771 gggggaaaagcccaaccatcgtcttcaacgaagctcctctcgaacggcagtcgggggaalc 830
Db 860 AGGAGGCAAGAAATCCGTGCATCGTGTGTGCAGATGCTGACTTGGACTTGGCCGTCGAGTG 919
QY 831 ggcaaaaggatttctcaaaaattcggggcaaaatttgggtccccccctcctgtttgctagtgc 890
Db 920 TGCTCACAGGGAGTGTCTTCAACCAAGGTCAGTGTGTACAGCGGCTCCAGGGTGT 979
QY 891 atggggaaaatttagcggagaaaaattccatggagtcctgcgtcatggctcatttggaggctgc- 949
Db 980 CGTGAAGAGCAGGTC-----TACGGGGAGTGTGTGAGGAGAGTGTGGAGTTCCGCCA 1032
QY 950 agagatggcttggccagaacccattggaacccaagaggacgcattggctcccttcgctcgaca 1009
Db 1033 AGAAGAGGCGGTTGGAGACCCCTTCGATGCCAAAACGGAGCAGGGGCTCAGATCGACC 1092
QY 1010 agtcccagtcacagagcttctgggttaacattgacgttgg---caaggataccgcgcagc 1066
Db 1093 AAAAGCAGTTTGACAAAATCCTCGAGCTGATTGAGAGTGGGAAGAGGAAGGGGCCAAGC 1152
QY 1067 tcctcactggcgttggtagaaaaggggcacaagggattcgcgattgaacgcgacgatattg 1126
Db 1153 TAGAATGTGGGGGTCAGCCATGGAGGACAGAGGGCTGTTTCATCAAAACCACGGTCTTCT 1212
QY 1127 tcaatcccaaacaccaggcagcaaaaatttgggttgaggagatccttggcccccgctctgtcca 1186
Db 1213 CAGATGTTACGGACAACATGAGGATTTGCCAAAGAGGAGATTTTCGGACCCAGTCCAGCCGA 1272
QY 1187 ttaagacgttcaagacggaagaagagggccattgagattgccaatgacacgacttatgggc 1246
Db 1273 TCCTGAAGTTCAAAACCCTGGAGGAGGTGATCAAGAGAGCGGAATAGCACCGACTATGGAC 1332
QY 1247 tagcctcggtcatttataccaaaatctcacaagggtctccgtgtctcgtcggcgctcg 1306
Db 1333 TCACAGCAGCAGTGTTCACCAAAAACCTGGACAAAAGCACTGAAGCTGGCTGGCTGCGCTCG 1392
QY 1307 agaccggtggcgtctcgatcaacttcccccttatcccccgagacacaaaactccgttttggcg 1366
Db 1393 AGTCGGGGACAGTCTGGATCAACTGCTACAAATGCAATTTATGCACAGGCTCCATTTGGTG 1452
QY 1367 gcatgaaacaatcgggctcaggcagagagctaggcgaagaaggcctcaaggcgctacttgg 1426
Db 1453 GCTTCAAAATGCTGGGAATGGCAGAGAACTAGGAGAATATGCTCTGGCTGAATATACAG 1512
QY 1427 agcccaagaccatlaatatatccacgtcaa 1454

Db 1513 AAGTGAAAACCGTCAACCATCAAACTCGA 1540
Search completed: April 26, 2002, 23:04:52
Job time: 13281 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2002, 02:58:36 ; Search time 4999.95 Seconds
(without alignments)
3951.948 Million cell updates/sec

Title: US-09-882-694A-4
Perfect score: 1464
Sequence: 1 atggttttttcgcctgacga.....tccacgtcaacatagagtga 1464

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0
Searched: 13736207 seqs, 6748477542 residues

Word size : 0
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_estl:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	22	1.5	566	9 AU204674	AU204674 AU204674
2	21	1.4	378	12 AQ848277	AQ848277 LMAJFV1_1
3	21	1.4	524	10 BG557867	BG557867 EM1_55_E1
4	21	1.4	757	9 AI068883	AI068883 mgae0004c
5	21	1.4	769	10 BI225854	BI225854 602950228
6	21	1.4	775	10 BI856603	BI856603 603385735
7	21	1.4	801	9 AI068982	AI068982 mgae0004d
8	21	1.4	852	10 BI258465	BI258465 602972456
9	21	1.4	930	10 BF035820	BF035820 601458344
10	21	1.4	938	9 AL536113	AL536113 AL536113
11	21	1.4	1243	10 BI457385	BI457385 603185515
12	20	1.4	534	10 BG816128	BG816128 dac07h02.
13	20	1.4	572	9 AI293860	AI293860 LP07108.3
14	20	1.4	582	10 BJ060784	BJ060784 BJ060784
15	20	1.4	583	9 AI259976	AI259976 LP03541.5
16	20	1.4	583	10 BI942019	BI942019 se82c09.y
17	20	1.4	591	10 BG905010	BG905010 TaLr1136E

18	1.4	595	9 AI261062	AI261062 LP05133.5
19	1.4	597	9 AI114304	AI114304 GH11111.3
20	1.4	604	9 AI296186	AI296186 LP10021.5
21	1.4	630	9 AI455414	AI455414 LD23985.3
22	1.4	677	10 BE282264	BE282264 601101239
23	1.3	115	10 BJ112053	BJ112053 BJ112053
24	1.3	188	9 AW230296	AW230296 up28g10.y
25	1.3	201	10 BJ121003	BJ121003 BJ121003
26	1.3	203	9 AV248967	AV248967 AV248967
27	1.3	207	10 BE439065	BE439065 CDO269.WH
28	1.3	209	10 BM090714	BM090714 ig17h05.y
29	1.3	218	9 BB117397	BB117397 BB117397
30	1.3	237	9 BB565455	BB565455 BB565455
31	1.3	252	10 BJ101954	BJ101954 BJ101954
32	1.3	261	9 AA197346	AA197346 zp92f10.s
33	1.3	308	10 M75758	M75758 CEL01A9S1 C
34	1.3	310	9 BB377860	BB377860 BB377860
35	1.3	358	9 AU109836	AU109836 AU109836
36	1.3	360	10 C49496	C49496 C49496 Yuji
37	1.3	361	9 AU109727	AU109727 AU109727
38	1.3	367	9 AW897385	AW897385 RC4-NN005
39	1.3	369	10 BE439116	BE439116 CDO269.WH
40	1.3	388	10 H79241	H79241 Yu27g08.r1
41	1.3	401	9 AW847618	AW847618 IL3-CT021
42	1.3	401	9 AW847619	AW847619 IL3-CT021
43	1.3	401	9 AW847629	AW847629 IL3-CT021
44	1.3	401	9 AW847634	AW847634 IL3-CT021
45	1.3	413	9 AW847540	AW847540 IL3-CT021

ALIGNMENTS

RESULT 1
AU204674
LOCUS AU204674 566 bp mRNA linear EST 17-JUL-2001
DEFINITION AU204674 unpublished oligo-capped cDNA library, stage L4
Caenorhabditis elegans cDNA clone yk840d11 5', mRNA sequence.
ACCESSION AU204674
VERSION AU204674.1 GI:14836273
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 566)
AUTHORS Kohara,Y., Shin-i,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.
and Sugano,S.
TITLE A complementary view of the C.elegans genome
JOURNAL Unpublished (2001)
COMMENT Contact: Yuji Kohara
Genome Biology Lab.
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
Location/Qualifiers
1. .566
/organism="Caenorhabditis elegans"
/strain="N2"
/db_xref="taxon:6239"
/clone="yk840d11"
/clone_lib="unpublished oligo-capped cDNA library, stage L4"
/sex="Hermaphrodite"
/tissue_type="whole animal"
/dev_stage="L4"

FEATURES
source

BASE COUNT 160 a 129 c 129 g 148 t
ORIGIN

Query Match 1.5%; Score 22; DB 9; Length 566;

Best Local Similarity 100.0%; Pred. No. 4.9;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 tcttcacaaatgaattcgt 56
|||||
Db 118 TCTTCATCAACAATGAATTCGT 139

RESULT 2
AQ848277/c
LOCUS
DEFINITION
LMAJFV1_lm59a07.x1 Leishmania major FV1 random genomic library
Leishmania major genomic clone LMAJFV1_lm59a07 3' similar to
SW:DHAM_LEITA Q25417 ALDEHYDE DEHYDROGENASE, MITOCHONDRIAL
PRECURSOR ;, DNA sequence.

ACCESSION
AQ848277
VERSION
AQ848277.1 GI:6052886
KEYWORDS
GSS.
SOURCE
Leishmania major.
ORGANISM
Leishmania major
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Leishmania.

REFERENCE
1 (bases 1 to 378)
AUTHORS
Akopyants,N.S., Clifton,S.W., Martin,J., Pape,D., Wylie,T., Li,L.,
Kissinger,J., Roos,D.S., Marra,M., Hillier,L., Chinwalla,A.,
Blistain,A., Schmitt,A., Person,B., Theising,B., Ritter,E., Ronko
,I., Bennett,J., Cole,R., Underwood,K., Cardenas,M., Gibbons,M.,
Harvey,N., McCann,R., Tsagarelshvili,R., Williams,T., Jackson,Y.,
Bowers,Y., Swaller,T., Waterston,R., Wilson,R. and Beverley,S.M.
A survey of the Leishmania major Friedlin strain V1 genome by
shotgun sequencing: a resource for DNA microarrays and expression
profiling

TITLE
shotgun sequencing: a resource for DNA microarrays and expression
profiling
JOURNAL
Mol. Biochem. Parasitol. 113 (2), 337-340 (2001)
MEDLINE
21192569
COMMENT
Contact: Akopyants, NS / Beverley, SM
WashU Leishmania Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

Library construction: Natalia S. Akopyants, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
If using this information please cite:
N.S. Akopyants and S.M. Beverley 'A survey of the Leishmania major
Friedlin strain V1 genome by shotgun sequencing' and the Washington
University Genome Sequencing Center For information on obtaining
clone material please contact: Natalia S. Akopyants Ph.D.
(natalia@borcim.wustl.edu) and/or Stephen M. Beverley Ph.D.
(beverley@borcim.wustl.edu)

Seq primer: -40UP from Gibco
Class: shotgun
High quality sequence stop: 82.

FEATURES
source
1..378
/organism="Leishmania major"
/strain="Friedlin strain V1"
/db_xref="taxon:5664"
/clone="LMAJFV1_lm59a07"
/clone_lib="Leishmania major FV1 random genomic library"
/lab_host="TOP10 (Invitrogen)"
/note="Vector: pZero-2 (Invitrogen); Site_1: EcoRV;
Genomic DNA was isolated from stationary phase cells. For
this library, DNA was sheared to give a tight size
distribution of 1-1.5kb fragments, blunt-ended with T4 DNA
polymerase, dephosphorylated with Shrimp Alkaline
Phosphatase and ligated into pZero-2 vector's EcoRV site."

BASE COUNT 77 a 117 c 104 g 80 t
ORIGIN

Query Match 1.4%; Score 21; DB 12; Length 378;
Best Local Similarity 100.0%; Pred. No. 15;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1159 gaggagatctttgccccgtc 1179
|||||
Db 306 GAGGAGATCTTTGGCCCCGTC 286

RESULT 3
BG557867
LOCUS
DEFINITION
EM1_55_E11.b1_A002 Embryo 1 (EM1) Sorghum bicolor cDNA, mRNA
sequence.
ACCESSION
BG557867
VERSION
BG557867.1 GI:13586865
KEYWORDS
EST.
SOURCE
sorghum.
ORGANISM
Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.
1 (bases 1 to 524)
AUTHORS
Reid,S.P., Cordonnier-Pratt,M.-M., Gingle,A. and Pratt,L.H.
TITLE
An EST database from Sorghum: developing embryos
JOURNAL
Unpublished (2000)
COMMENT
Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmpratt@uga.edu

Sequences have been trimmed to exclude POLYA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: JEN REV
High quality sequence stop: 491
POLYA=No.

FEATURES
source
1..524
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Embryo 1 (EM1)"
/note="Organ: Embryos germinated for 24 hr; Vector:
pBluescript II from lambda Zap II; Site_1: XhoI; Site_2:
EcoRI; The library was made from poly-A RNA in the cloning
vector lambda ZAP II. Clones to be sequenced were
prepared by mass excision."
BASE COUNT 68 a 206 c 155 g 95 t
ORIGIN

Query Match 1.4%; Score 21; DB 10; Length 524;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 505 gccccgcacctgcgcgcgcgc 525
|||||
Db 30 GCCCCCGCCCTCGCCGCCGCGC 50

RESULT 4
AI068883
LOCUS
DEFINITION
mgae0004cD07f Magnaporthe grisea Appressorium Stage cDNA Library
Magnaporthe grisea cDNA clone mgae0004cD07f 5', mRNA sequence.
ACCESSION
AI068883
VERSION
AI068883.1 GI:3391858
KEYWORDS
EST.
SOURCE
Magnaporthe grisea.
ORGANISM
Magnaporthe grisea
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariomycetes incertae sedis; Magnaportheaceae; Magnaporthe.
1 (bases 1 to 757)

BASE COUNT 68 a 206 c 155 g 95 t
ORIGIN

Query Match 1.4%; Score 21; DB 10; Length 524;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

FEATURES
source
1..524
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Embryo 1 (EM1)"
/note="Organ: Embryos germinated for 24 hr; Vector:
pBluescript II from lambda Zap II; Site_1: XhoI; Site_2:
EcoRI; The library was made from poly-A RNA in the cloning
vector lambda ZAP II. Clones to be sequenced were
prepared by mass excision."
BASE COUNT 68 a 206 c 155 g 95 t
ORIGIN

Query Match 1.4%; Score 21; DB 12; Length 378;
Best Local Similarity 100.0%; Pred. No. 15;

AUTHORS Choi,W., Fang,E., Sasinowski,M., Wing,R. and Dean,R.A.
TITLE Expressed sequence characterization during appressorium formation
in rice blast fungus, Magnaporthe grisea
JOURNAL Unpublished (1998)
COMMENT Contact: Dean,R.A.
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson University, Clemson, SC 29634, USA
Tel: 864 656 5737
Fax: 864 656 4293
Email: rdean@clemson.edu
Seq primer: T3 primer (AATTAAACCCCTCACTAAAGGG)
High quality sequence stop: 325.
Location/Qualifiers

FEATURES source
1..757
/organism="Magnaporthe grisea"
/strain="70-15"
/db_xref="taxon:148305"
/clone="mgae0004cd07f"
/clone_lib="Magnaporthe grisea Appressorium Stage cDNA Library"
/dev_stage="Germinated conidia on appressorium-inductive surface"
/note="Vector: pBluescriptII SK(+) Vector; Site_1: EcoRI; Site_2: XhoI; The appressorium formation-specific cDNA library was constructed from conidia germinated for 5-8 hr on an inductive surface. The library has an average insert size of 1.5 kbp."
BASE COUNT 181 a 218 c 196 g 161 t 1 others
ORIGIN

Query Match 1.4%; Score 21; DB 9; Length 757;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1159 gaggagatctttggcccgctc 1179
|||||
Db 431 GAGGAGATCTTTGGCCCCGTC 451

RESULT 5
BI225854/c
LOCUS
DEFINITION 602950228F1 NIH_MGC_8 Homo sapiens cDNA clone IMAGE:5094559 5',
mRNA sequence.
ACCESSION BI225854
VERSION BI225854.1 GI:14679298
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 769)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1863 row: g column: 08
High quality sequence stop: 723.
Location/Qualifiers

FEATURES source
1..769
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5094559"

/clone_lib="NIH_MGC_8"
/tissue_type="Burkitt lymphoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lymph; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using 2AP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 182 a 208 c 223 g 156 t
ORIGIN

Query Match 1.4%; Score 21; DB 10; Length 769;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 895 ggaatttagcggagaaattc 915
|||||
Db 168 GGAAATTAGCGGAGAAATTC 148

RESULT 6
BI856603/c
LOCUS
DEFINITION 603385735F1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:5394441 5',
mRNA sequence.
ACCESSION BI856603
VERSION BI856603.1 GI:15997350
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 775)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTP
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12005 row: f column: 10
High quality sequence stop: 508.
Location/Qualifiers

FEATURES source
1..775
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5394441"
/clone_lib="NIH_MGC_87"
/tissue_type="mammary adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: breast; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.383 kb. Library enriched for full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 197 a 201 c 217 g 160 t
ORIGIN

Query Match 1.4%; Score 21; DB 10; Length 775;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 895 ggaatttagcggagaaattc 915

/note="Organ: ovary; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.8 kb. Library constructed by Life
Technologies."

BASE COUNT 222 a 295 c 251 g 162 t
ORIGIN

Query Match 1.4%; Score 21; DB 10; Length 930;

Best Local Similarity 100.0%; Pred. No. 17;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 895 ggaaatttagcggagaaattc 915

|||||

Db 125 GCAAAATTTAGCGGAGAAATTC 105

RESULT 10

AL536113

DEFINITION AL536113 LTI_FL013_FBrn1 Homo sapiens cDNA clone CS0DF02YK12 5 prime, mRNA sequence.

ACCESSION

AL536113

VERSION AL536113.1 GI:12799606

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheraia; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 938)

Li,W.B., Gruber,C., Jesse,J. and Polayes,D.

Full-length cDNA libraries and normalization

Unpublished (2001)

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

Location/Qualifiers

1. .938

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="CS0DF02YK12"

/clone_lib="LTI_FL013_FBrn1"

/dev_stage="pooled tissue from post conception fetuses (20

week, 24 week and 26 week)"

/lab_host="DH10B"

/note="Organ: Fetal brain; Vector: pCMVSPORT 6; 1st strand

cDNA was primed with a NotI-oligo(dT) primer. Five prime

end enriched, double-stranded cDNA was digested with Not I

and cloned into the Not I and Eco RV sites of the

pCMVSPORT 6 vector. Library was constructed by Life

Technologies. Contact : Feng Liang Life Technologies, a

division of Invitrogen 9800 Medical Center Drive Rockville

, Maryland 20850, USA Fax : (1) 301 610 8371 Email :

fliang@lifetech.com URL :

http://fulllength.invitrogen.com"

0 a 526 c 291 g 120 t 1 others

BASE COUNT

ORIGIN

Query Match 1.4%; Score 21; DB 9; Length 938;

Best Local Similarity 100.0%; Pred. No. 17;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 505 gccccgcctcgccgcgcgc 525

|||||

Db 263 GCCCGCGCCTCGCGCGGC 283

RESULT 11

BI457385/c

LOCUS

DEFINITION 603185515F1 NIH_MGC_42 Homo sapiens cDNA clone IMAGE:5258306 5',

ACCESSION

BI457385

VERSION

BI457385.1

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheraia; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1243)

NIH-MGC http://mgs.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone Distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLCM1869 row: n column: 03

High quality sequence stop: 263.

Location/Qualifiers

1. .1243

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:5258306"

/clone_lib="NIH_MGC_42"

/tissue_type="epithelioid carcinoma cell line"

/lab_host="DH10B (phage-resistant)"

/note="Organ: pancreas; Vector: pOTB7; Site_1: XhoI;

Site_2: EcoRI; cDNA made by oligo-dT priming.

Directionally cloned into EcoRI/XhoI sites using the

following 5' adaptor: GGCACGAG(G). Size-selected >500bp

for average insert size 1.8kb. Library constructed by Ling

Hong in the laboratory of Gerald M. Rubin (University of

California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies).

Note: this is a NIH_MGC Library. |"

379 a 355 c 318 g 191 t

BASE COUNT

ORIGIN

Query Match 1.4%; Score 21; DB 10; Length 1243;

Best Local Similarity 100.0%; Pred. No. 18;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 895 ggaaatttagcggagaaattc 915

|||||

Db 214 GCAAAATTTAGCGGAGAAATTC 194

RESULT 12

BG816128/c

LOCUS

DEFINITION

BG816128

dac07h02.y1 NICHD XGC Emb2 Xenopus laevis cDNA clone IMAGE:4406618

5' similar to SW:OXYB_HUMAN P22059 OXYSTEROL-BINDING PROTEIN. ;,

mrna sequence.

BG816128

VERSION

BG816128.1

KEYWORDS

EST.

SOURCE

African clawed frog.

ORGANISM

Xenopus laevis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;

Xenopodinae; Xenopus.

1 (bases 1 to 534)

Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.

, Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person

,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,

Waterston,R. and Willson,R.

Washu Xenopus EST project, 1999

TITLE

JOURNAL
COMMENT

Unpublished (1999)
Other_ESTs: dac07h02.x1
Contact: Sandy Clifton, ph.D.
WashU xenopus EST project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Library constructed by Life Technologies. DNA Sequencing by:
Washington University Genome Sequencing Center
Clone distribution: xenopus clones from this library are available
through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
High quality sequence stop: 434.

FEATURES

source

1. .534
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="IMAGE:4406618"
/clone_lib="NICHD XGC Emb2"
/tissue_type="embryo, stage 17/19"
/lab_host="DH10B (phage-resistant)"
/note="Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI;
Cloned unidirectionally. Primer: Oligo dT. Average insert
size 2.1 kb. Constructed by Life Technologies."

BASE COUNT
ORIGIN

158 a 102 c 129 g 145 t
Query Match 1.4%; Score 20; DB 10; Length 534;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 524 gctgctcttcattcttcaaa 543
|||||
Db 190 GCTGCTCCTTCATCTTCAAA 171

RESULT 13
AI293860/C
LOCUS

DEFINITION AI293860 572 bp mRNA linear EST 19-APR-2001
LP07108.3prime LP Drosophila melanogaster larval-early pupal POT2
Drosophila melanogaster cDNA clone LP07108 3prime, mRNA sequence.

ACCESSION

AI293860

VERSION

AI293860.1 GI:3943267

KEYWORDS

EST.

SOURCE

fruit fly.

ORGANISM

Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

1 (bases 1 to 572)

AUTHORS

Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
Lewis,S. and Rubin,G.M.

TITLE

BDGP/HM1 Drosophila EST Project

JOURNAL

Unpublished (2001)

COMMENT

Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: <http://www.fruitfly.org/EST>, est@fruitfly.berkeley.edu
Based upon the presence of a xhoI site followed by a run of 14 or
more T residues at the beginning of the sequence, this clone
probably contains an inverted insert. The sequence has been trimmed
and the T residues removed.
Plate: 71 row: A column: 8
High quality sequence stop: 467.

FEATURES

source

1. .572
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="LP07108"

/clone_lib="LP Drosophila melanogaster larval-early pupal
POT2"
/sex="male and female"
/dev_stage="larvae-pupae"
/lab_host="DH5-alpha"
/note="Organ: whole body; Vector: POT2; Site_1: EcoRI;
Site_2: XhoI; Sized fractionated cDNAs were directly
ligated into POT2. Plasmid cDNA library."

BASE COUNT 151 a 154 c 130 g 137 t
ORIGIN

Query Match 1.4%; Score 20; DB 9; Length 572;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1159 gaggagatcttggcccgct 1178
|||||
Db 402 GAGGAGATCTTTGGCCCGCT 383

RESULT 14

BJ060784/C

LOCUS

DEFINITION

BJ060784 NIBB Mochii normalized xenopus tailbud library xenopus

laevis cDNA clone XLO66n04 5', mRNA sequence.

ACCESSION BJ060784.1 GI:17500525

VERSION BJ060784

KEYWORDS EST.

SOURCE African clawed frog.

ORGANISM Xenopus laevis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;

Xenopodinae; Xenopus.

1 (bases 1 to 582)

REFERENCE Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-I,T. and Kohara

,Y.

TITLE Expressed genes in X. laevis embryo

JOURNAL Unpublished (2001)

COMMENT Contact: Tadasu Shin-I

Center For Genetic Resource Information

National Institute of Genetics

1111 Yata, Mishima, Shizuoka 411-8540, Japan

Tel: 81-559-81-6856

Fax: 81-559-81-6855

Email: tshini@genes.nig.ac.jp.

FEATURES Location/Qualifiers

1. .582

/organism="Xenopus laevis"

/db_xref="taxon:8355"

/clone="XLO66n04"

/clone_lib="NIBB Mochii normalized Xenopus tailbud

library"

/tissue_type="whole embryo"

/dev_stage="stage 25"

BASE COUNT 171 a 115 c 124 g 170 t 2 others

ORIGIN

Query Match 1.4%; Score 20; DB 10; Length 582;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 524 gctgctcttcattcttcaaa 543

|||||

Db 68 GCTGCTCCTTCATCTTCAAA 49

RESULT 15

AI259976

LOCUS

DEFINITION

LP03541.5prime LP Drosophila melanogaster larval-early pupal POT2

583 bp mRNA linear EST 19-APR-2001

Drosophila melanogaster cDNA clone LP03541 5prime, mRNA sequence.

ACCESSION

AI259976

VERSION

AI259976.1 GI:3867501

KEYWORDS

EST

SOURCE

fruit fly.

ORGANISM

Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 583)
Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,
Lewis, S. and Rubin, G.M.
BDGP/HMI Drosophila EST project
Unpublished (2001)
Contact: Stapleton, M.
BDGP

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: <http://www.fruitfly.org/EST>, est@fruitfly.berkeley.edu
Sequence is the complete cDNA insert.
Plate: 35 row: D column: 5
High quality sequence stop: 581.

FEATURES

source

1..583

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/clone="LP03541"

/clone_lib="LP Drosophila melanogaster larval-early pupal

POT2"

/sex="male and female"

/dev_stage="larvae-pupae"

/lab_host="DH5-alpha"

/note="Organ: whole body; Vector: POT2; Site_1: EcoRI;

Site_2: XhoI; Sized fractionated cDNAs were directly

ligated into POT2. Plasmid cDNA library."

BASE COUNT

146 a 131 c 153 g 153 t

ORIGIN

Query Match

Best Local Similarity 1.4%; Score 20; DB 9; Length 583;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1159 gaggagatctttggcccggt 1178

|||||

Db 172 GAGGAGATCTTTGGCCCGGT 191

Search completed: April 27, 2002, 02:58:44

Job time: 13436 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic ~ nucleic search, using sw model

Run on: April 27, 2002, 07:43:52 ; Search time 95.74 Seconds
(without alignments)
2809.045 Million cell updates/sec

Title: US-09-882-694A-4
Perfect score: 1464
Sequence: 1 atggttcttcgcctgacga.....tccacgtcaacatagagtga 1464.

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 139433 seqs, 91850399 residues

Word size : 0

Total number of hits satisfying chosen parameters: 278866

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Listing first 45 summaries

Database : Pending_Patents_NA_New:*
1: /cgn2_6/ptodata/1/pna/pct_NEW_COMB.seq.*
2: /cgn2_6/ptodata/1/pna/US06_NEW_COMB.seq.*
3: /cgn2_6/ptodata/1/pna/US07_NEW_COMB.seq.*
4: /cgn2_6/ptodata/1/pna/US08_NEW_COMB.seq.*
5: /cgn2_6/ptodata/1/pna/US09_NEW_COMB.seq.*
6: /cgn2_6/ptodata/1/pna/US10_NEW_COMB.seq.*
7: /cgn2_6/ptodata/1/pna/US60_NEW_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	18	1.2	33965	6 US-10-105-299-12544	Sequence 12544, A
c 2	17	1.2	259	5 US-09-975-254-13294	Sequence 13294, A
3	17	1.2	569	6 US-10-102-524-206	Sequence 206, App
4	17	1.2	961	6 US-10-105-299-11655	Sequence 11655, A
5	17	1.2	964	6 US-10-105-299-2100	Sequence 2100, App
6	17	1.2	3725	6 US-10-116-802-334	Sequence 334, App
c 7	17	1.2	32493	5 US-09-539-698A-6	Sequence 6, Appli
8	17	1.2	39339	6 US-10-105-299-7119	Sequence 7119, Ap
9	16	1.1	99	6 US-10-106-698-3137	Sequence 3137, Ap
10	16	1.1	197	6 US-10-097-105-1475	Sequence 1475, Ap
11	16	1.1	218	5 US-09-540-210B-3171	Sequence 3171, Ap
12	16	1.1	235	5 US-09-540-210B-16680	Sequence 16680, A
13	16	1.1	246	5 US-09-975-254-13104	Sequence 13104, A
14	16	1.1	248	5 US-09-975-254-18042	Sequence 18042, A
c 15	16	1.1	251	5 US-09-540-210B-205	Sequence 205, App
c 16	16	1.1	412	7 US-60-365-384-462	Sequence 462, App
c 17	16	1.1	412	7 US-60-365-384-463	Sequence 463, App
18	16	1.1	807	6 US-10-107-096-1	Sequence 1, Appli
19	16	1.1	978	5 US-09-540-209B-4283	Sequence 4283, Ap
20	16	1.1	1139	6 US-10-102-806-122	Sequence 122, App
c 21	16	1.1	1335	5 US-09-312-283B-76	Sequence 76, Appl
c 22	16	1.1	1335	5 US-09-312-283B-261	Sequence 261, App
23	16	1.1	1516	7 US-60-365-384-120	Sequence 120, App
24	16	1.1	1576	7 US-60-365-384-121	Sequence 121, App
25	16	1.1	1609	5 US-09-896-908-1	Sequence 1, Appli
26	16	1.1	1647	5 US-09-896-908-14	Sequence 14, Appl

27	16	1.1	2061	5 US-09-540-209B-817	Sequence 817, App
28	16	1.1	2898	6 US-10-029-397A-52	Sequence 52, Appl
29	16	1.1	2900	6 US-10-029-397A-55	Sequence 55, Appl
30	16	1.1	2901	6 US-10-029-397A-57	Sequence 57, Appl
31	16	1.1	2902	6 US-10-029-397A-56	Sequence 56, Appl
32	16	1.1	2912	6 US-10-029-397A-51	Sequence 51, Appl
33	16	1.1	2922	6 US-10-029-397A-50	Sequence 50, Appl
34	16	1.1	2923	6 US-10-029-397A-54	Sequence 54, Appl
35	16	1.1	2932	6 US-10-029-397A-53	Sequence 53, Appl
36	16	1.1	5196	5 US-09-630-630B-28	Sequence 28, Appl
c 37	16	1.1	7025	6 US-10-105-299-9628	Sequence 9628, Ap
c 38	16	1.1	9525	6 US-10-105-299-8161	Sequence 8161, Ap
c 39	16	1.1	9525	6 US-10-105-299-8162	Sequence 8162, Ap
c 40	16	1.1	9525	6 US-10-105-299-8172	Sequence 8172, Ap
c 41	16	1.1	9525	6 US-10-105-299-8173	Sequence 8173, Ap
c 42	16	1.1	11538	6 US-10-105-299-12735	Sequence 12735, A
c 43	16	1.1	12390	6 US-10-105-299-8164	Sequence 8164, Ap
c 44	16	1.1	12390	6 US-10-105-299-8175	Sequence 8175, Ap
45	16	1.1	13003	6 US-10-112-699-2911	Sequence 2911, Ap

ALIGNMENTS

RESULT 1
US-10-105-299-12544/c
; Sequence 12544, Application US/10105299
; GENERAL INFORMATION:
; APPLICANT: Rosen, et. al
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS950
; CURRENT APPLICATION NUMBER: US/10/105,299
; CURRENT FILING DATE: 2002-03-26
; NUMBER OF SEQ ID NOS: 15197
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12544
; LENGTH: 33965
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-105-299-12544

Query Match 1.2%; Score 18; DB 6; Length 33965;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 529 tccttcacatcttcaaaagcc 546
|||||
Db 13262 TCCTTCATCTTCAAAGCC 13245

RESULT 2
US-09-975-254-13294/c
; Sequence 13294, Application US/09975254
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: Heck, Gregory R.
; APPLICANT: La Rosa, Thomas J.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(15309)B
; CURRENT APPLICATION NUMBER: US/09/975,254
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US/09/263,191
; PRIOR FILING DATE: 1999-03-05
; NUMBER OF SEQ ID NOS: 31255
; SEQ ID NO 13294
; LENGTH: 259
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: 700907504H1
US-09-975-254-13294

```

Query Match      1.2%; Score 17; DB 5; Length 259;
Best Local Similarity 100.0%; Pred. No. 9;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1135 aaaccaggcagcaaaat 1151
Db 75 AAACCAGGCAGCAAAAT 59

RESULT 3
US-10-102-524-206
; Sequence 206, Application US/101052524
; GENERAL INFORMATION:
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Gaiger, Alexander
; APPLICANT: Gordon, Brian
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; FILE OF INVENTION: THERAPY AND DIAGNOSIS OF KIDNEY CANCER
; FILE REFERENCE: 210121.572
; CURRENT APPLICATION NUMBER: US/10/102,524
; CURRENT FILING DATE: 2002-03-19
; NUMBER OF SEQ ID NOS: 1863
; SOFTWARE: FastSeq for Windows version 4.0
; SEQ ID NO 206
; LENGTH: 569
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-102-524-206

Query Match      1.2%; Score 17; DB 6; Length 569;
Best Local Similarity 100.0%; Pred. No. 9.3;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 641 gagtgcgggtgaagca 657
Db 229 gagtgcgggtgaagca 245

RESULT 4
US-10-105-299-11655
; Sequence 11655, Application US/10105299
; GENERAL INFORMATION:
; APPLICANT: Rosen, et. al
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS950
; CURRENT APPLICATION NUMBER: US/10/105,299
; CURRENT FILING DATE: 2002-03-26
; NUMBER OF SEQ ID NOS: 15197
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11655
; LENGTH: 961
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-105-299-11655

Query Match      1.2%; Score 17; DB 6; Length 961;
Best Local Similarity 100.0%; Pred. No. 9.4;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 118 actgatgttcacgtggc 134
Db 323 actgatgttcacgtggc 339

RESULT 5
US-10-105-299-2100
```

```

; Sequence 2100, Application US/10105299
; GENERAL INFORMATION:
; APPLICANT: Rosen, et. al
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS950
; CURRENT APPLICATION NUMBER: US/10/105,299
; CURRENT FILING DATE: 2002-03-26
; NUMBER OF SEQ ID NOS: 15197
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2100
; LENGTH: 964
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (883)..(883)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-105-299-2100

Query Match      1.2%; Score 17; DB 6; Length 964;
Best Local Similarity 100.0%; Pred. No. 9.4;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 118 actgatgttcacgtggc 134
Db 321 actgatgttcacgtggc 337

RESULT 6
US-10-116-802-334
; Sequence 334, Application US/10116802
; GENERAL INFORMATION:
; APPLICANT: Amy Lasek
; TITLE OF INVENTION: GENES EXPRESSED IN LUNG CANCER
; FILE REFERENCE: PA-0045 US
; CURRENT APPLICATION NUMBER: US/10/116,802
; CURRENT FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: 60/281,593
; PRIOR FILING DATE: 2001-04-04
; NUMBER OF SEQ ID NOS: 519
; SOFTWARE: PERL Program
; SEQ ID NO 334
; LENGTH: 3725
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 474926.11
US-10-116-802-334

Query Match      1.2%; Score 17; DB 6; Length 3725;
Best Local Similarity 100.0%; Pred. No. 9.9;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 530 ccttcacatcttcaaagcc 546
Db 167 ccttcacatcttcaaagcc 183

RESULT 7
US-09-539-698A-5/c
; Sequence 6, Application US/09539698A
; GENERAL INFORMATION:
; APPLICANT: GenStar Therapeutics Corporation
; APPLICANT: Alemany, Ramon
; APPLICANT: Fang, Xiangming
; APPLICANT: Zhang, Wei-wei
; APPLICANT: Robert, Sobol
; TITLE OF INVENTION: Complementary-Adenoviral vector System
; FILE REFERENCE: 97-087-B
```

```
; CURRENT APPLICATION NUMBER: US/09/539,698A
; CURRENT FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 08/797,160
; PRIOR FILING DATE: 1997-02-10
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 32493
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(32493)
; OTHER INFORMATION: GTV8053 plasmid vector
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (10461)..(16291)
; OTHER INFORMATION: complement (10461..16291), PSA promoter/enhancer sequence
; FEATURE:
; NAME/KEY: polyA_signal
; LOCATION: (7276)..(7423)
; OTHER INFORMATION: polyA signal sequence
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (7441)..(10440)
; OTHER INFORMATION: ElaElb sequence coding sequence in 10440..7441 orientation
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (16742)..(17138)
; OTHER INFORMATION: RSV LTR
; FEATURE:
; NAME/KEY: Intron
; LOCATION: (17164)..(17284)
; OTHER INFORMATION: SV40 intron
; FEATURE:
; NAME/KEY: gene
; LOCATION: (17395)..(18262)
; OTHER INFORMATION: human B7-1 cDNA sequence
; FEATURE:
; NAME/KEY: polyA_signal
; LOCATION: (18292)..(18490)
; OTHER INFORMATION: SV40 poly A signal sequence
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (18523)..(19785)
; OTHER INFORMATION: EF promoter sequence
; FEATURE:
; NAME/KEY: gene
; LOCATION: (19805)..(20305)
; OTHER INFORMATION: human interferon gamma (hIFNg) cDNA sequence
; FEATURE:
; NAME/KEY: polyA_signal
; LOCATION: (20712)..(21113)
; OTHER INFORMATION: polyA signal sequence
; FEATURE:
; NAME/KEY: gene
; LOCATION: (23662)..(24559)
; OTHER INFORMATION: EGFP sequence
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (24567)..(32050)
; OTHER INFORMATION: AFP DE3 sequence
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (487)..(7251)
; OTHER INFORMATION: Alb UE5 sequence
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (22957)..(23646)
; OTHER INFORMATION: SV40 promoter and first intron
; FEATURE:
; NAME/KEY: LTR
; LOCATION: (1)..(100)
```

```
; OTHER INFORMATION: adenovirus (Ad) ITR5'
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (236)..(373)
; OTHER INFORMATION: packaging signal
; FEATURE:
; NAME/KEY: LTR
; LOCATION: (32393)..(32493)
; OTHER INFORMATION: adenovirus (Ad) ITR3'
US-09-539-698A-6

Query Match          1.2%; Score 17; DB 5; Length 32493;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 188 agggcccatggaagaag 204
      |||||
Db 20668 AGGCCCATGGAGAAG 20652

RESULT 8
US-10-105-299-7119
; Sequence 7119, Application US/10105299
; GENERAL INFORMATION:
; APPLICANT: Rosen, et. al
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS950
; CURRENT APPLICATION NUMBER: US/10/105,299
; CURRENT FILING DATE: 2002-03-26
; NUMBER OF SEQ ID NOS: 15197
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7119
; LENGTH: 39339
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-105-299-7119

Query Match          1.2%; Score 17; DB 6; Length 39339;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 599 ccggattccctcctgga 615
      |||||
Db 36 ccggattccctcctgga 52

RESULT 9
US-10-106-698-3137
; Sequence 3137, Application US/10106698
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypep.
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 3137
; LENGTH: 99
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (7)..(7)
```


; PRIOR FILING DATE: February 13, 1997
; PRIOR APPLICATION NUMBER: 09/035,172
; PRIOR FILING DATE: March 4, 1998
; PRIOR APPLICATION NUMBER: 60/040,431
; PRIOR FILING DATE: March 5, 1997
; PRIOR APPLICATION NUMBER: 09/041,894
; PRIOR FILING DATE: March 12, 1998
; PRIOR APPLICATION NUMBER: 60/040,199
; PRIOR FILING DATE: March 14, 1997
; PRIOR APPLICATION NUMBER: 09/050,817
; PRIOR FILING DATE: March 30, 1998
; PRIOR APPLICATION NUMBER: 60/043,792
; PRIOR FILING DATE: April 11, 1997
; PRIOR APPLICATION NUMBER: 09/074,999
; PRIOR FILING DATE: May 8, 1998
; PRIOR APPLICATION NUMBER: 60/048,431
; PRIOR FILING DATE: May 29, 1997
; PRIOR APPLICATION NUMBER: 09/107,592
; PRIOR FILING DATE: June 30, 1998
; PRIOR APPLICATION NUMBER: 60/052,751
; PRIOR FILING DATE: July 1, 1997
; PRIOR APPLICATION NUMBER: 09/094,079
; PRIOR FILING DATE: June 9, 1998
; PRIOR APPLICATION NUMBER: 60/049,975
; PRIOR FILING DATE: June 13, 1997
; NUMBER OF SEQ ID NOS: 35654
; SOFTWARE: PERL Program
; SEQ ID NO 3171
; LENGTH: 218
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: hu00348038
US-09-540-210B-3171

Query Match 1.1%; Score 16; DB 5; Length 218;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 733 acactcaagtccaaca 748
| | | | | | | | | | | | | | | | | | | | | |
Db 72 acactcaagtccaaca 87

RESULT 12
US-09-540-210B-16680
; Sequence 16680, Application US/09540210B
; GENERAL INFORMATION:
; APPLICANT: Seilhamer, Jeffrey J.
; APPLICANT: Delegeane, Angelo M.
; APPLICANT: Stuart, Susan G.
; APPLICANT: Stuve, Laura L.
; APPLICANT: Mullahy, Sara J.
; APPLICANT: Naughton, Rebecca E.
; TITLE OF INVENTION: POLYNUCLEOTIDES OF URINARY TRACT TISSUE
; FILE REFERENCE: PD-1037 CIP
; CURRENT APPLICATION NUMBER: US/09/540,210B
; CURRENT FILING DATE: 2002-04-03
; PRIOR APPLICATION NUMBER: 08/972,899
; PRIOR FILING DATE: November 18, 1997
; PRIOR APPLICATION NUMBER: 08/395,244
; PRIOR FILING DATE: February 27, 1995
; PRIOR APPLICATION NUMBER: 08/722,922
; PRIOR FILING DATE: September 27, 1996
; PRIOR APPLICATION NUMBER: 60/005,526
; PRIOR FILING DATE: September 29, 1995
; PRIOR APPLICATION NUMBER: 08/824,029
; PRIOR FILING DATE: March 25, 1997
; PRIOR APPLICATION NUMBER: 60/014,010
; PRIOR FILING DATE: March 25, 1996
; PRIOR APPLICATION NUMBER: 08/826,847

; PRIOR FILING DATE: April 10, 1997
; PRIOR APPLICATION NUMBER: 60/015,533
; PRIOR FILING DATE: April 10, 1996
; PRIOR APPLICATION NUMBER: 08/903,555
; PRIOR FILING DATE: July 31, 1997
; PRIOR APPLICATION NUMBER: 60/023,308
; PRIOR FILING DATE: July 31, 1996
; PRIOR APPLICATION NUMBER: 08/862,178
; PRIOR FILING DATE: May 22, 1997
; PRIOR APPLICATION NUMBER: 60/018,217
; PRIOR FILING DATE: May 23, 1996
; PRIOR APPLICATION NUMBER: 08/881,589
; PRIOR FILING DATE: June 24, 1997
; PRIOR APPLICATION NUMBER: 60/021,275
; PRIOR FILING DATE: June 25, 1996
; PRIOR APPLICATION NUMBER: 08/903,802
; PRIOR FILING DATE: July 31, 1997
; PRIOR APPLICATION NUMBER: 60/023,308
; PRIOR FILING DATE: July 31, 1996
; PRIOR APPLICATION NUMBER: 08/905,881
; PRIOR FILING DATE: August 1, 1997
; PRIOR APPLICATION NUMBER: 60/025,204
; PRIOR FILING DATE: August 1, 1996
; PRIOR APPLICATION NUMBER: 08/903,471
; PRIOR FILING DATE: July 30, 1997
; PRIOR APPLICATION NUMBER: 60/025,478
; PRIOR FILING DATE: July 31, 1996
; PRIOR APPLICATION NUMBER: 08/903,556
; PRIOR FILING DATE: July 31, 1997
; PRIOR APPLICATION NUMBER: 60/025,217
; PRIOR FILING DATE: August 22, 1996
; PRIOR APPLICATION NUMBER: 08/937,142
; PRIOR FILING DATE: September 23, 1997
; PRIOR APPLICATION NUMBER: 60/026,598
; PRIOR FILING DATE: September 24, 1996
; PRIOR APPLICATION NUMBER: 08/960,746
; PRIOR FILING DATE: October 29, 1997
; PRIOR APPLICATION NUMBER: 60/030,144
; PRIOR FILING DATE: October 30, 1996
; PRIOR APPLICATION NUMBER: 08/826,847
; PRIOR FILING DATE: April 10, 1997
; PRIOR APPLICATION NUMBER: 60/015,533
; PRIOR FILING DATE: April 10, 1996
; PRIOR APPLICATION NUMBER: 08/755,524
; PRIOR FILING DATE: November 22, 1996
; PRIOR APPLICATION NUMBER: 60/007,495
; PRIOR FILING DATE: November 22, 1995
; PRIOR APPLICATION NUMBER: 09/021,031
; PRIOR FILING DATE: February 10, 1998
; PRIOR APPLICATION NUMBER: 60/039,325
; PRIOR FILING DATE: February 13, 1997
; PRIOR APPLICATION NUMBER: 09/035,172
; PRIOR FILING DATE: March 4, 1998
; PRIOR APPLICATION NUMBER: 60/040,431
; PRIOR FILING DATE: March 5, 1997
; PRIOR APPLICATION NUMBER: 09/041,894
; PRIOR FILING DATE: March 12, 1998
; PRIOR APPLICATION NUMBER: 60/040,199
; PRIOR FILING DATE: March 14, 1997
; PRIOR APPLICATION NUMBER: 09/050,817
; PRIOR FILING DATE: March 30, 1998
; PRIOR APPLICATION NUMBER: 60/043,792
; PRIOR FILING DATE: April 11, 1997
; PRIOR APPLICATION NUMBER: 09/074,999
; PRIOR FILING DATE: May 8, 1998
; PRIOR APPLICATION NUMBER: 60/048,431
; PRIOR FILING DATE: May 29, 1997
; PRIOR APPLICATION NUMBER: 09/107,592
; PRIOR FILING DATE: June 30, 1998
; PRIOR APPLICATION NUMBER: 60/052,751
; PRIOR FILING DATE: July 1, 1997
; PRIOR APPLICATION NUMBER: 09/094,079
; PRIOR FILING DATE: June 9, 1998


```
; PRIOR APPLICATION NUMBER: 60/049,975
; PRIOR FILING DATE: June 13, 1997
; NUMBER OF SEQ ID NOS: 35654
; SOFTWARE: PERL Program
; SEQ ID NO 16680
; LENGTH: 235
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: hu00457807
; NAME/KEY: unsure
; LOCATION: 108, 115
; OTHER INFORMATION: a, t, c, g, or other
US-09-540-210B-16680
```

```
Query Match      1.1%; Score 16; DB 5; Length 235;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 733 acactcaagtccaaca 748
    |||||
Db 23 acactcaagtccaaca 38
```

RESULT 13

```
US-09-975-254-13104
; Sequence 13104, Application US/09975254
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: Heck, Gregory R.
; APPLICANT: La Rosa, Thomas J.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(15309)B
; CURRENT APPLICATION NUMBER: US/09/975,254
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US/09/263,191
; PRIOR FILING DATE: 1999-03-05
; NUMBER OF SEQ ID NOS: 31255
; SEQ ID NO 13104
; LENGTH: 246
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: 700907183H1
US-09-975-254-13104
```

```
Query Match      1.1%; Score 16; DB 5; Length 246;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 908 agaaattccatggagt 923
    |||||
Db 169 agaaattccatggagt 184
```

RESULT 14

```
US-09-975-254-18042
; Sequence 18042, Application US/09975254
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: Heck, Gregory R.
; APPLICANT: La Rosa, Thomas J.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(15309)B
; CURRENT APPLICATION NUMBER: US/09/975,254
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US/09/263,191
; PRIOR FILING DATE: 1999-03-05
; NUMBER OF SEQ ID NOS: 31255
```

```
; SEQ ID NO 18042
; LENGTH: 248
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: 700955062H1
US-09-975-254-18042
```

```
Query Match      1.1%; Score 16; DB 5; Length 248;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 908 agaaattccatggagt 923
    |||||
Db 184 agaaattccatggagt 199
```

RESULT 15

```
US-09-540-210B-205/c
; Sequence 205, Application US/09540210B
; GENERAL INFORMATION:
; APPLICANT: Seilhamer, Jeffrey J.
; APPLICANT: Delegeane, Angelo M.
; APPLICANT: Stuart, Susan G.
; APPLICANT: Stuve, Laura L.
; APPLICANT: Mullahy, Sara J.
; APPLICANT: Naughton, Rebecca E.
; TITLE OF INVENTION: POLYNUCLEOTIDES OF URINARY TRACT TISSUE
; FILE REFERENCE: PD-1037 CIP
; CURRENT APPLICATION NUMBER: US/09/540,210B
; CURRENT FILING DATE: 2002-04-03
; PRIOR APPLICATION NUMBER: 08/972,899
; PRIOR FILING DATE: November 18, 1997
; PRIOR APPLICATION NUMBER: 08/395,244
; PRIOR FILING DATE: February 27, 1995
; PRIOR APPLICATION NUMBER: 08/722,922
; PRIOR FILING DATE: September 27, 1996
; PRIOR APPLICATION NUMBER: 60/005,526
; PRIOR FILING DATE: September 29, 1995
; PRIOR APPLICATION NUMBER: 08/824,029
; PRIOR FILING DATE: March 25, 1997
; PRIOR APPLICATION NUMBER: 60/014,010
; PRIOR FILING DATE: March 25, 1996
; PRIOR APPLICATION NUMBER: 08/826,847
; PRIOR FILING DATE: April 10, 1997
; PRIOR APPLICATION NUMBER: 60/015,533
; PRIOR FILING DATE: April 10, 1996
; PRIOR APPLICATION NUMBER: 08/903,555
; PRIOR FILING DATE: July 31, 1997
; PRIOR APPLICATION NUMBER: 60/023,308
; PRIOR FILING DATE: July 31, 1996
; PRIOR APPLICATION NUMBER: 08/862,178
; PRIOR FILING DATE: May 22, 1997
; PRIOR APPLICATION NUMBER: 60/018,217
; PRIOR FILING DATE: May 23, 1996
; PRIOR APPLICATION NUMBER: 08/881,589
; PRIOR FILING DATE: June 24, 1997
; PRIOR APPLICATION NUMBER: 60/021,275
; PRIOR FILING DATE: June 25, 1996
; PRIOR APPLICATION NUMBER: 08/903,802
; PRIOR FILING DATE: July 31, 1997
; PRIOR APPLICATION NUMBER: 60/023,308
; PRIOR FILING DATE: July 31, 1996
; PRIOR APPLICATION NUMBER: 08/905,881
; PRIOR FILING DATE: August 1, 1997
; PRIOR APPLICATION NUMBER: 60/025,204
; PRIOR FILING DATE: August 1, 1996
; PRIOR APPLICATION NUMBER: 08/903,471
; PRIOR FILING DATE: July 30, 1997
; PRIOR APPLICATION NUMBER: 60/025,478
; PRIOR FILING DATE: July 31, 1996
; PRIOR APPLICATION NUMBER: 08/903,556
; PRIOR FILING DATE: July 31, 1997
```

; PRIOR APPLICATION NUMBER: 60/025,217
; PRIOR FILING DATE: August 22, 1996
; PRIOR APPLICATION NUMBER: 08/937,142
; PRIOR FILING DATE: September 23, 1997
; PRIOR APPLICATION NUMBER: 60/026,598
; PRIOR FILING DATE: September 24, 1996
; PRIOR APPLICATION NUMBER: 08/960,746
; PRIOR FILING DATE: October 29, 1997
; PRIOR APPLICATION NUMBER: 60/030,144
; PRIOR FILING DATE: October 30, 1996
; PRIOR APPLICATION NUMBER: 08/826,847
; PRIOR FILING DATE: April 10, 1997
; PRIOR APPLICATION NUMBER: 60/015,533
; PRIOR FILING DATE: April 10, 1996
; PRIOR APPLICATION NUMBER: 08/755,524
; PRIOR FILING DATE: November 22, 1996
; PRIOR APPLICATION NUMBER: 60/007,495
; PRIOR FILING DATE: November 22, 1995
; PRIOR APPLICATION NUMBER: 09/021,031
; PRIOR FILING DATE: February 10, 1998
; PRIOR APPLICATION NUMBER: 60/039,325
; PRIOR FILING DATE: February 13, 1997
; PRIOR APPLICATION NUMBER: 09/035,172
; PRIOR FILING DATE: March 4, 1998
; PRIOR APPLICATION NUMBER: 60/040,431
; PRIOR FILING DATE: March 5, 1997
; PRIOR APPLICATION NUMBER: 09/041,894
; PRIOR FILING DATE: March 12, 1998
; PRIOR APPLICATION NUMBER: 60/040,199
; PRIOR FILING DATE: March 14, 1997
; PRIOR APPLICATION NUMBER: 09/050,817
; PRIOR FILING DATE: March 30, 1998
; PRIOR APPLICATION NUMBER: 60/043,792
; PRIOR FILING DATE: April 11, 1997
; PRIOR APPLICATION NUMBER: 09/074,999
; PRIOR FILING DATE: May 8, 1998
; PRIOR APPLICATION NUMBER: 60/048,431
; PRIOR FILING DATE: May 29, 1997
; PRIOR APPLICATION NUMBER: 09/107,592
; PRIOR FILING DATE: June 30, 1998
; PRIOR APPLICATION NUMBER: 60/052,751
; PRIOR FILING DATE: July 1, 1997
; PRIOR APPLICATION NUMBER: 09/094,079
; PRIOR FILING DATE: June 9, 1998
; PRIOR APPLICATION NUMBER: 60/049,975
; PRIOR FILING DATE: June 13, 1997
; NUMBER OF SEQ ID NOS: 35654
; SOFTWARE: PERL Program
; SEQ ID NO 205
; LENGTH: 251
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: hu00875909
; NAME/KEY: unsure
; LOCATION: 203
; OTHER INFORMATION: a, t, c, g, or other
US-09-540-210B-205

Query Match 1.1%; Score 16; DB 5; Length 251;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 472 aacgcgacttttcttt 487
Db 157 AACGCGACTTTTCTTT 142

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2002, 07:41:43 ; Search time 8172.53 Seconds
(without alignments)
3875.604 Million cell updates/sec

Title: US-09-882-694A-4
Perfect score: 1464
Sequence: 1 atgggtcttgcctgacga.....tccacgtcaacatagagtga 1464

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 21979536 seqs, 10817449327 residues

Word size : 0
Total number of hits satisfying chosen parameters: 43959072

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending_Patents_NA_Main:*

1:	/cgn2_6/ptodata/2/pna/PCTUS_COMB.seq:*
2:	/cgn2_6/ptodata/2/pna/US06_COMB.seq:*
3:	/cgn2_6/ptodata/2/pna/US07_COMB.seq:*
4:	/cgn2_6/ptodata/2/pna/US080_COMB.seq:*
5:	/cgn2_6/ptodata/2/pna/US081_COMB.seq:*
6:	/cgn2_6/ptodata/2/pna/US082_COMB.seq:*
7:	/cgn2_6/ptodata/2/pna/US083_COMB.seq:*
8:	/cgn2_6/ptodata/2/pna/US084_COMB.seq:*
9:	/cgn2_6/ptodata/2/pna/US085_COMB.seq:*
10:	/cgn2_6/ptodata/2/pna/US086_COMB.seq:*
11:	/cgn2_6/ptodata/2/pna/US087_COMB.seq:*
12:	/cgn2_6/ptodata/2/pna/US088_COMB.seq:*
13:	/cgn2_6/ptodata/2/pna/US089_COMB.seq:*
14:	/cgn2_6/ptodata/2/pna/US090_COMB.seq:*
15:	/cgn2_6/ptodata/2/pna/US091_COMB.seq:*
16:	/cgn2_6/ptodata/2/pna/US092_COMB.seq:*
17:	/cgn2_6/ptodata/2/pna/US093_COMB.seq:*
18:	/cgn2_6/ptodata/2/pna/US094_COMB.seq:*
19:	/cgn2_6/ptodata/2/pna/US095A_COMB.seq:*
20:	/cgn2_6/ptodata/2/pna/US095B_COMB.seq:*
21:	/cgn2_6/ptodata/2/pna/US095C_COMB.seq:*
22:	/cgn2_6/ptodata/2/pna/US095D_COMB.seq:*
23:	/cgn2_6/ptodata/2/pna/US096A_COMB.seq:*
24:	/cgn2_6/ptodata/2/pna/US096B_COMB.seq:*
25:	/cgn2_6/ptodata/2/pna/US096C_COMB.seq:*
26:	/cgn2_6/ptodata/2/pna/US096D_COMB.seq:*
27:	/cgn2_6/ptodata/2/pna/US096E_COMB.seq:*
28:	/cgn2_6/ptodata/2/pna/US097A_COMB.seq:*
29:	/cgn2_6/ptodata/2/pna/US097B_COMB.seq:*
30:	/cgn2_6/ptodata/2/pna/US097C_COMB.seq:*
31:	/cgn2_6/ptodata/2/pna/US098A_COMB.seq:*
32:	/cgn2_6/ptodata/2/pna/US098B_COMB.seq:*
33:	/cgn2_6/ptodata/2/pna/US098C_COMB.seq:*
34:	/cgn2_6/ptodata/2/pna/US099A_COMB.seq:*
35:	/cgn2_6/ptodata/2/pna/US099B_COMB.seq:*
36:	/cgn2_6/ptodata/2/pna/US099C_COMB.seq:*
37:	/cgn2_6/ptodata/2/pna/US100_COMB.seq:*
38:	/cgn2_6/ptodata/2/pna/US101_COMB.seq:*
39:	/cgn2_6/ptodata/2/pna/US6000_COMB.seq:*
40:	/cgn2_6/ptodata/2/pna/US6001_COMB.seq:*
41:	/cgn2_6/ptodata/2/pna/US6002_COMB.seq:*
42:	/cgn2_6/ptodata/2/pna/US6003_COMB.seq:*
43:	/cgn2_6/ptodata/2/pna/US6004_COMB.seq:*

44:	/cgn2_6/ptodata/2/pna/US6005_COMB.seq:*
45:	/cgn2_6/ptodata/2/pna/US6006_COMB.seq:*
46:	/cgn2_6/ptodata/2/pna/US6007_COMB.seq:*
47:	/cgn2_6/ptodata/2/pna/US6008_COMB.seq:*
48:	/cgn2_6/ptodata/2/pna/US6009_COMB.seq:*
49:	/cgn2_6/ptodata/2/pna/US6010_COMB.seq:*
50:	/cgn2_6/ptodata/2/pna/US6011_COMB.seq:*
51:	/cgn2_6/ptodata/2/pna/US6012_COMB.seq:*
52:	/cgn2_6/ptodata/2/pna/US6013_COMB.seq:*
53:	/cgn2_6/ptodata/2/pna/US6014_COMB.seq:*
54:	/cgn2_6/ptodata/2/pna/US6015_COMB.seq:*
55:	/cgn2_6/ptodata/2/pna/US6016_COMB.seq:*
56:	/cgn2_6/ptodata/2/pna/US6017_COMB.seq:*
57:	/cgn2_6/ptodata/2/pna/US6018_COMB.seq:*
58:	/cgn2_6/ptodata/2/pna/US6019_COMB.seq:*
59:	/cgn2_6/ptodata/2/pna/US6020_COMB.seq:*
60:	/cgn2_6/ptodata/2/pna/US6021_COMB.seq:*
61:	/cgn2_6/ptodata/2/pna/US6022_COMB.seq:*
62:	/cgn2_6/ptodata/2/pna/US6023_COMB.seq:*
63:	/cgn2_6/ptodata/2/pna/US6024_COMB.seq:*
64:	/cgn2_6/ptodata/2/pna/US6025_COMB.seq:*
65:	/cgn2_6/ptodata/2/pna/US6026_COMB.seq:*
66:	/cgn2_6/ptodata/2/pna/US6027_COMB.seq:*
67:	/cgn2_6/ptodata/2/pna/US6028_COMB.seq:*
68:	/cgn2_6/ptodata/2/pna/US6029_COMB.seq:*
69:	/cgn2_6/ptodata/2/pna/US6030_COMB.seq:*
70:	/cgn2_6/ptodata/2/pna/US6031_COMB.seq:*
71:	/cgn2_6/ptodata/2/pna/US6032_COMB.seq:*
72:	/cgn2_6/ptodata/2/pna/US6033_COMB.seq:*
73:	/cgn2_6/ptodata/2/pna/US6034_COMB.seq:*
74:	/cgn2_6/ptodata/2/pna/US6035_COMB.seq:*
75:	/cgn2_6/ptodata/2/pna/US6036_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	score	% Query Match	Length	DB	ID	Description
1	1464	100.0	1464	17	US-09-351-224-4	Sequence 4, Appli
2	1464	100.0	1464	17	US-09-351-224E-4	Sequence 4, Appli
3	1464	100.0	1464	17	US-09-351-823-4	Sequence 4, Appli
4	1464	100.0	1464	26	US-09-677-488-4	Sequence 4, Appli
5	1464	100.0	1464	26	US-09-677-488A-4	Sequence 4, Appli
6	1464	100.0	1464	26	US-09-677-682-4	Sequence 4, Appli
7	1464	100.0	1464	26	US-09-677-682A-4	Sequence 4, Appli
8	1464	100.0	1464	26	US-09-677-682B-4	Sequence 4, Appli
9	1464	100.0	1464	33	US-09-882-694-4	Sequence 4, Appli
10	1464	100.0	1464	33	US-09-882-694A-4	Sequence 4, Appli
11	22	1.5	12332	24	US-09-620-392-55184	Sequence 55184, A
12	22	1.5	20140	28	US-09-702-134-4827	Sequence 4827, Ap
13	22	1.5	20140	31	US-09-815-264-82296	Sequence 82296, A
14	21	1.4	340	31	US-09-804-730-13078	Sequence 13078, A
15	21	1.4	340	57	US-60-189-657-13072	Sequence 13072, A
16	21	1.4	363	26	US-09-675-784A-1380	Sequence 1380, Ap
17	21	1.4	471	23	US-09-605-702-16631	Sequence 16631, A
18	21	1.4	538	23	US-09-605-702-16630	Sequence 16630, A
19	21	1.4	648	19	US-09-526-263A-51	Sequence 51, Appl
20	21	1.4	657	26	US-09-675-784A-6649	Sequence 6649, Ap
21	21	1.4	704	16	US-09-270-767-11385	Sequence 11385, A
22	21	1.4	1068	18	US-09-417-507-12226	Sequence 12226, A
23	21	1.4	1134	18	US-09-417-507-12222	Sequence 12222, A
24	21	1.4	3103	24	US-09-620-392-66768	Sequence 66768, A
25	21	1.4	28326	24	US-09-620-392-67807	Sequence 67807, A
26	21	1.4	28326	28	US-09-702-134-9191	Sequence 9191, Ap
27	21	1.4	28326	31	US-09-815-264-63345	Sequence 63345, A
28	21	1.4	97086	64	US-60-258-279-8	Sequence 8, Appli
29	20	1.4	361	64	US-60-253-654-4231	Sequence 4231, Ap
30	20	1.4	361	64	US-60-255-592-4231	Sequence 4231, Ap
31	20	1.4	378	26	US-09-667-188A-6695	Sequence 6695, Ap

[illegible]

; OTHER INFORMATION: aldehyde dehydrogenase , coding sequence
US-09-351-823-4

Query Match		100.0%;	Score 1464;	DB 17;	Length 1464;
Best Local Similarity		100.0%;	Pred. No. 0;		
Matches 1464;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	atgggtctcttcgctgacgaatacaagagtggaactcttcatcaacaatgaattcgtctcc	60		
Db	1	atgggtctcttcgctgacgaatacaagagtggaactcttcatcaacaatgaattcgtctcc	60		
QY	61	tccaaggggtccgagagattaaacgctcacgaacccgtgggacgaatccaccgttgccact	120		
Db	61	tccaaggggtccgagagattaaacgctcacgaacccgtgggacgaatccaccgttgccact	120		
QY	121	gatgttcacgtggccaaacgctgacagatgctgacagtgacagtgacgcttcggtcagggcg	180		
Db	121	gatgttcacgtggccaaacgctgacagatgctgacagtgacagtgacgcttcggtcagggcg	180		
QY	181	gtcaaaaaggcccatggaagaagttcacaggtgcacaacgcgcgcgtgcattgcttaag	240		
Db	181	gtcaaaaaggcccatggaagaagttcacaggtgcacaacgcgcgcgtgcattgcttaag	240		
QY	241	ttcgcggacctgcgcgagaaacgcgcgagaagctcgtcgtcttgagtcgctgcccacc	300		
Db	241	ttcgcggacctgcgcgagaaacgcgcgagaagctcgtcgtcttgagtcgctgcccacc	300		
QY	301	ggtagacgggtgctcgatgacatcactcatttcgacattccaaacatggtctcgtgttcgc	360		
Db	301	ggtagacgggtgctcgatgacatcactcatttcgacattccaaacatggtctcgtgttcgc	360		
QY	361	tactatgcagggctgggcccgaacagatgcgcggaaagacctttcccgaggacaaaggcaag	420		
Db	361	tactatgcagggctgggcccgaacagatgcgcggaaagacctttcccgaggacaaaggcaag	420		
QY	421	ccgaattggcgttacgagccgatgggggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	480		
Db	421	ccgaattggcgttacgagccgatgggggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	480		
QY	481	tttctttacgtcggctggaagatagccccgcctcgcgcgcgcgcgcgcgcgcgcgcgcgc	540		
Db	481	tttctttacgtcggctggaagatagccccgcctcgcgcgcgcgcgcgcgcgcgcgcgcgc	540		
QY	541	aaagcctcggagaaatcccgctgggctgtctgggctcgcgcgcgcgcgcgcgcgcgcgcgc	600		
Db	541	aaagcctcggagaaatcccgctgggctgtctgggctcgcgcgcgcgcgcgcgcgcgcgcgc	600		
QY	601	ggattccctcgtgagtcgtgagttcctcactggagacagatgacgggtgaagcattg	660		
Db	601	ggattccctcgtgagtcgtgagttcctcactggagacagatgacgggtgaagcattg	660		
QY	661	gcgtgcacatggacattgcgaagatcagcttcacaaagatctgtcgggtggcgcgcgc	720		
Db	661	gcgtgcacatggacattgcgaagatcagcttcacaaagatctgtcgggtggcgcgcgc	720		
QY	721	gtcaagcaagcaaacactcaactcaacatgaagcgcgtcactctagaaactgggggaaaaag	780		
Db	721	gtcaagcaagcaaacactcaactcaacatgaagcgcgtcactctagaaactgggggaaaaag	780		
QY	781	ccaaccatcgtcttcaacgaagctcctcgcgaacgagtcgggggaatcggcaaggat	840		
Db	781	ccaaccatcgtcttcaacgaagctcctcgcgaacgagtcgggggaatcggcaaggat	840		
QY	841	ttctcaaaattcgggcaaaatttgggtcccccctcctgtttgtagtgcaatggggaat	900		
Db	841	ttctcaaaattcgggcaaaatttgggtcccccctcctgtttgtagtgcaatggggaat	900		
QY	901	ttagcggagaaaattccatggagtcctcgtcatttggagctgtcagagatggctt	960		
Db	901	ttagcggagaaaattccatggagtcctcgtcatttggagctgtcagagatggctt	960		
QY	961	ggcagaacccattggaaccacagaggacgcatggtcccttcgtcgaacaagtccagtag	1020		

Db	961	ggcagaacccattggaaccacagaggacgcatggtcccttcgtcgaacaagtcacgtac	1020
QY	1021	gacagagctcttggttaacattgacgcttggaagagataccgcgacgtcctcactggcgtt	1080
Db	1021	gacagagctcttggttaacattgacgcttggaagagataccgcgacgtcctcactggcgtt	1080
QY	1081	ggtagaagggcgacaaggattcgcgattgaacgacgacatattgtcaatcccaacca	1140
Db	1081	ggtagaagggcgacaaggattcgcgattgaacgacgacatattgtcaatcccaacca	1140
QY	1141	ggcagcaaaatttggtttgaggagatctttggcccccgtcttgcattaaagacgttcaag	1200
Db	1141	ggcagcaaaatttggtttgaggagatctttggcccccgtcttgcattaaagacgttcaag	1200
QY	1201	acggaagaaggccattgagattgccaatgacaacgacttatgggctagcctcgttcatt	1260
Db	1201	acggaagaaggccattgagattgccaatgacaacgacttatgggctagcctcgttcatt	1260
QY	1261	tataccaaatctcacaagggtgtccgtgtctcgtgtctcgtcgcgcgcgcgcgcgcgcgc	1320
Db	1261	tataccaaatctcacaagggtgtccgtgtctcgtgtctcgtcgcgcgcgcgcgcgcgcgc	1320
QY	1321	tcgatcaactcccttttatccccgagacacaaactccgtttggcggcgcgcgcgcgcgcgc	1380
Db	1321	tcgatcaactcccttttatccccgagacacaaactccgtttggcggcgcgcgcgcgcgcgc	1380
QY	1381	ggctcaggcagagagtaggaggaagggtcctcgtgtctcgtcgcgcgcgcgcgcgcgcgc	1440
Db	1381	ggctcaggcagagagtaggaggaagggtcctcgtgtctcgtcgcgcgcgcgcgcgcgcgc	1440
QY	1441	aatatccacgtcaacatagagtgga	1464
Db	1441	aatatccacgtcaacatagagtgga	1464

RESULT 4

US-09-677-488-4
; Sequence 4, Application US/09677488
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jon
; TITLE OF INVENTION: Compositions and Methods for Fummonisin Detoxification
; FILE REFERENCE: 35718/204100 (5718-111A)
; CURRENT APPLICATION NUMBER: US/09/677,488
; CURRENT FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1464
; TYPE: DNA
; ORGANISM: Exophiala spinifera
; FEATURE:
; OTHER INFORMATION: aldehyde dehydrogenase , fully spliced cDNA
US-09-677-488-4

Query Match		100.0%;	Score 1464;	DB 26;	Length 1464;
Best Local Similarity		100.0%;	Pred. No. 0;		
Matches 1464;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	atgggtctcttcgctgacgaatacaagagtggaactcttcatcaacaatgaattcgtctcc	60		
Db	1	atgggtctcttcgctgacgaatacaagagtggaactcttcatcaacaatgaattcgtctcc	60		
QY	61	tccaaggggtccgagagattaaacgctcacgaacccgtgggacgaatccaccgttgccact	120		
Db	61	tccaaggggtccgagagattaaacgctcacgaacccgtgggacgaatccaccgttgccact	120		
QY	121	gatgttcacgtggccaaacgc	180		
Db	121	gatgttcacgtggccaaacgc	180		

QY 361 tactatgcaggctgggcccgaacaagatcccgaaagacacctttcccagagacaacggcgaag 420
Db 361 tactatgcaggctgggcccgaacaagatcccgaaagacacctttcccagagacaacggcgaag 420

QY 421 ccgaattggcgttaacgagccgatgggggtgtgtgctggtattgccagctggaacgcgact 480
Db 421 ccgaattggcgttaacgagccgatgggggtgtgtgctggtattgccagctggaacgcgact 480

QY 481 ttctttacgtcggctggaagatagccccccgcctgcgcggctgctcttcattcttc 540
Db 481 ttctttacgtcggctggaagatagccccccgcctgcgcggctgctcttcattcttc 540

QY 541 aaagcctcggagaaatccccgcctggggttctgggctcgcctcgtctctcttcgcagaagcc 600
Db 541 aaagcctcggagaaatccccgcctggggttctgggctcgcctcgtctctcttcgcagaagcc 600

QY 601 ggattccctcctggagtcgtgcagttcctcacttgagcacgagtgacgggtgaagcattg 660
Db 601 ggattccctcctggagtcgtgcagttcctcacttgagcacgagtgacgggtgaagcattg 660

QY 661 gcgtcgcacatggacattgcgaagatcagcttcacaagatctgtcggcgtggccgcgccc 720
Db 661 gcgtcgcacatggacattgcgaagatcagcttcacaagatctgtcggcgtggccgcgccc 720

QY 721 gtcaagcaagcaacactcaagtccaacatgaagcgcgtcactctagaaactgggggaaag 780
Db 721 gtcaagcaagcaacactcaagtccaacatgaagcgcgtcactctagaaactgggggaaag 780

QY 781 ccaaccatcgtcttcaacgaagtcctctcgaacggcagtcggtggggaatcggaagat 840
Db 781 ccaaccatcgtcttcaacgaagtcctctcgaacggcagtcggtggggaatcggaagat 840

QY 841 ttctcaaaattcgggcaaaatttgggtcccccctcctgtttgctagtgcgaatgggaaat 900
Db 841 ttctcaaaattcgggcaaaatttgggtcccccctcctgtttgctagtgcgaatgggaaat 900

QY 901 ttacgggagaaattccatggagtcctcgtcatttggaggtgtcagagatggcctt 960
Db 901 ttacgggagaaattccatggagtcctcgtcatttggaggtgtcagagatggcctt 960

QY 961 ggccagaacccattggaacccaaagaggacgcatgggtcccttcgtcgacaagtcccagtac 1020
Db 961 ggccagaacccattggaacccaaagaggacgcatgggtcccttcgtcgacaagtcccagtac 1020

QY 1021 gacagagtcttgggtaaacattgacgttggcaaggatataccgcgcagctcctcactggcgtt 1080
Db 1021 gacagagtcttgggtaaacattgacgttggcaaggatataccgcgcagctcctcactggcgtt 1080

QY 1081 ggtagaaaggcgcaagggtatccgattgaacccgacgatatattgtcaatcccaaacca 1140
Db 1081 ggtagaaaggcgcaagggtatccgattgaacccgacgatatattgtcaatcccaaacca 1140

QY 1141 ggcagcaaaatttgggttaggagatctttggcccccgctcttgtccattaaagacgttcaag 1200
Db 1141 ggcagcaaaatttgggttaggagatctttggcccccgctcttgtccattaaagacgttcaag 1200

QY 1201 acggaagaagagccatttgagattgccaatgacacgacttatgggctagcctcgggtcatt 1260
Db 1201 acggaagaagagccatttgagattgccaatgacacgacttatgggctagcctcgggtcatt 1260

QY 1261 tataccaaatctcacaaggggtctccgtgtctcgtcggcgtcgcagacccggtggcgtc 1320
Db 1261 tataccaaatctcacaaggggtctccgtgtctcgtcggcgtcgcagacccggtggcgtc 1320

QY 1321 tcgatcaacttcccctttatccccgagacacaaactccgtttggcggcatgaaacaatcg 1380
Db 1321 tcgatcaacttcccctttatccccgagacacaaactccgtttggcggcatgaaacaatcg 1380

QY 1381 ggctcaggcagagagcttaggcgaagaagggtcgaaggcgtacttggagcccaagaccatt 1440
Db 1381 ggctcaggcagagagcttaggcgaagaagggtcgaaggcgtacttggagcccaagaccatt 1440

QY 1441 aatatccacgtcaacatagagtga 1464
Db 1441 aatatccacgtcaacatagagtga 1464

RESULT 6
US-09-677-682-4
; Sequence 4, Application US/09677682
; GENERAL INFORMATION:
; APPLICANT: DuVick, Jon
; TITLE OF INVENTION: Compositions and Methods for Fummonisin Detoxification
; FILE REFERENCE: 35718/204101 (5718-111B)
; CURRENT APPLICATION NUMBER: US/09/677,682
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1464
; TYPE: DNA
; ORGANISM: Exophiala spinifera
; FEATURE:
; OTHER INFORMATION: aldehyde dehydrogenase , fully spliced cDNA
US-09-677-682-4

Query Match 100.0%; Score 1464; DB 26; Length 1464;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1464; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggttcttcttcgctgacgaatacaagaagtgaactcttcatcaacaatgaattcgtctcc 60
Db 1 atggttcttcttcgctgacgaatacaagaagtgaactcttcatcaacaatgaattcgtctcc 60

QY 61 tccaaaggggtcgcagagagattaaacgtcacgaaccccgtaggacgaatccaccgttgcact 120
Db 61 tccaaaggggtcgcagagagattaaacgtcacgaaccccgtaggacgaatccaccgttgcact 120

QY 121 gatgttcacgtggcccaacgcggccgagtgtgcacagtgacgtagccgcttcggtgcaggcg 180
Db 121 gatgttcacgtggcccaacgcggccgagtgtgcacagtgacgtagccgcttcggtgcaggcg 180

QY 181 gtcaaaaaggcccatgggaagaatttcacaggttcacaaacgcgcggcggtgcattaaag 240
Db 181 gtcaaaaaggcccatgggaagaatttcacaggttcacaaacgcgcggcggtgcattaaag 240

QY 241 ttcgcggacctcgcgcgagaagaacgcgcgagaagctcgtcgtctggagtcgctgccacc 300
Db 241 ttcgcggacctcgcgcgagaagaacgcgcgagaagctcgtcgtctggagtcgctgccacc 300

QY 301 ggtagaccggtgtcgtatgatcactcatttccacattccaaaacatggtctcgtgttccgc 360
Db 301 ggtagaccggtgtcgtatgatcactcatttccacattccaaaacatggtctcgtgttccgc 360

QY 361 tactatgcaggctgggcccgaacaagatcgcgggaaagaccttccccgaggaacaacggcaag 420
Db 361 tactatgcaggctgggcccgaacaagatcgcgggaaagaccttccccgaggaacaacggcaag 420

QY 421 ccgaattggcgttacgagccgatgggggtgtgtgtgctggtattgccagctggaacgcgact 480
Db 421 ccgaattggcgttacgagccgatgggggtgtgtgtgctggtattgccagctggaacgcgact 480

QY 481 ttctttacgtcggctggaagatagccccccgcctcgcgcggctgctcttcattcttc 540
Db 481 ttctttacgtcggctggaagatagccccccgcctcgcgcggctgctcttcattcttc 540

QY 541 aaagcctcggagaaatccccgcctggggttctgggctcgcctcgtctcttcgcagaagcc 600
Db 541 aaagcctcggagaaatccccgcctggggttctgggctcgcctcgtctcttcgcagaagcc 600

QY 601 ggattccctcctggagtcgtgcagttcctcacttgacgagtgacgggtgaagcattg 660
Db 601 ggattccctcctggagtcgtgcagttcctcacttgacgagtgacgggtgaagcattg 660

Db 601 ggattccctcctggagtcgtgcagttcctcactggagcacgagtgacgggtgaagcattg 660
Qy 661 gcgtcgacatggacattgcgaagatcagcttcacaagatctgtcgcggtggccgcgc 720
Db 661 ggcgtcgacatggacattgcgaagatcagcttcacaagatctgtcgcggtggccgcgc 720
Qy 721 gtcaagcaagcaacactcaagtcacaacatgaagcgctcactctagaaactggggaaaag 780
Db 721 gtcaagcaagcaacactcaagtcacaacatgaagcgctcactctagaaactggggaaaag 780
Qy 781 ccaaccatcgcttctcaacgaagctcctctcgaaacggcagtcggggaaatcgcaaaagat 840
Db 781 ccaaccatcgcttctcaacgaagctcctctcgaaacggcagtcggggaaatcgcaaaagat 840
Qy 841 ttctcaaaattcgggcaaaatttgggtcccccctcctctgttgcagtcgaatgggaaaat 900
Db 841 ttctcaaaattcgggcaaaatttgggtcccccctcctctgttgcagtcgaatgggaaaat 900
Qy 901 ttagcggagaaaattccatggagtcgcgtcatggctcatttggaggtgtcagagatggctt 960
Db 901 ttagcggagaaaattccatggagtcgcgtcatggctcatttggaggtgtcagagatggctt 960
Qy 961 ggcacgaacccattggaaccccaagagacgcatggctcctctcgtcgacaagtccagtaac 1020
Db 961 ggcacgaacccattggaaccccaagagacgcatggctcctctcgtcgacaagtccagtaac 1020
Qy 1021 gacagagtccttgggttaacattgacgttggcaaggataccgcgcagctcctcactggcgtt 1080
Db 1021 gacagagtccttgggttaacattgacgttggcaaggataccgcgcagctcctcactggcgtt 1080
Qy 1081 ggtagaagggcgacaaaggattcgcgattgaaccgacgatatattgtcaatcccaaacca 1140
Db 1081 ggtagaagggcgacaaaggattcgcgattgaaccgacgatatattgtcaatcccaaacca 1140
Qy 1141 ggcagcaaaatttgggttggagagatcttggccccctgttgcattaaagcgttcaag 1200
Db 1141 ggcagcaaaatttgggttggagagatcttggccccctgttgcattaaagcgttcaag 1200
Qy 1201 acggaagaagagccattgagattgccaatgacacgacattatgggctagcctcggtcatt 1260
Db 1201 acggaagaagagccattgagattgccaatgacacgacattatgggctagcctcggtcatt 1260
Qy 1261 tataccaaatcttcaacaggggtctcgtgtctcgtgcgcgtcgagaccggtggcgtc 1320
Db 1261 tataccaaatcttcaacaggggtctcgtgtctcgtgcgcgtcgagaccggtggcgtc 1320
Qy 1321 tcgatcaacttcccttattatccccggagacacaaactccgttggcggcatgaaacaatcg 1380
Db 1321 tcgatcaacttcccttattatccccggagacacaaactccgttggcggcatgaaacaatcg 1380
Qy 1381 ggctcaggcagagagcttaggcgaagaagggtcctcaagcgctacttggagcccaagaccatt 1440
Db 1381 ggctcaggcagagagcttaggcgaagaagggtcctcaagcgctacttggagcccaagaccatt 1440
Qy 1441 aatatccacgtcaacatagagtga 1464
Db 1441 aatatccacgtcaacatagagtga 1464

RESULT 7
US-09-677-682A-4
; Sequence 4, Application US/09677682A
; GENERAL INFORMATION:
; APPLICANT: Duwick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; TITLE OF INVENTION: Detoxification
; FILE REFERENCE: 35718/204101
; CURRENT APPLICATION NUMBER: US/09/677,682A
; CURRENT FILING DATE: 2000-10-02

; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1464
; TYPE: DNA
; ORGANISM: Exophiala spinifera
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)..(0)
; OTHER INFORMATION: aldehyde dehydrogenase, fully spliced cDNA
US-09-677-682A-4

Query Match 100.0%; Score 1464; DB 26; Length 1464;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1464; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 atggttcttctgcctgacgaataacaaagatgaactcttccatcaacaatgaattcgtctcc 60
Db 1 atggttcttctgcctgacgaataacaaagatgaactcttccatcaacaatgaattcgtctcc 60
Qy 61 tccaaggggtccgagagattaaacgtcacgaacccgtgggacgaatccaccgttgcacct 120
Db 61 tccaaggggtccgagagattaaacgtcacgaacccgtgggacgaatccaccgttgcacct 120
Qy 121 gatgttcacgtggccaaacgcggccgattgtcgacagtgacagtgacagtgacagtgac 180
Db 121 gatgttcacgtggccaaacgcggccgattgtcgacagtgacagtgacagtgacagtgac 180
Qy 181 gtcaaaaaggcccatggaaagattcacaggtgcacaaacgcgcgcgtgcattcttaag 240
Db 181 gtcaaaaaggcccatggaaagattcacaggtgcacaaacgcgcgcgtgcattcttaag 240
Qy 241 ttgcgcgacctgcgcgagaaagacgcgcgagaaagctgcgtctgtgagtcgctgcccacc 300
Db 241 ttgcgcgacctgcgcgagaaagacgcgcgagaaagctgcgtctgtgagtcgctgcccacc 300
Qy 301 ggtagaccgggtgcgatgatcactcatttccgacattcccaaacatggtctcgtgttgcg 360
Db 301 ggtagaccgggtgcgatgatcactcatttccgacattcccaaacatggtctcgtgttgcg 360
Qy 361 tactatgcaggtggccgcgacagatcgccgaaagaccttcccgaggacaaacggcaag 420
Db 361 tactatgcaggtggccgcgacagatcgccgaaagaccttcccgaggacaaacggcaag 420
Qy 421 ccgaattggcgttacgacgcgatgggggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 480
Db 421 ccgaattggcgttacgacgcgatgggggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 480
Qy 481 ttcttttacgtcggtggaagatagccccccgcctcgcgcgcgcgcgcgcgcgcgcgcgcgc 540
Db 481 ttcttttacgtcggtggaagatagccccccgcctcgcgcgcgcgcgcgcgcgcgcgcgcgc 540
Qy 541 aaagcctcggagaaaatccccgcgtgggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 600
Db 541 aaagcctcggagaaaatccccgcgtgggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 600
Qy 601 ggattccctcctggagtcgtgcagttccctcactggagcacgagtgacgggtgaagcattg 660
Db 601 ggattccctcctggagtcgtgcagttccctcactggagcacgagtgacgggtgaagcattg 660
Qy 661 gcgtcgacatggacattgcgaagatcagcttcacaagatctgtcgcggtggccgcgcgc 720
Db 661 gcgtcgacatggacattgcgaagatcagcttcacaagatctgtcgcggtggccgcgcgc 720
Qy 721 gtcaagcaagcaacactcaagtcacaacatgaagcgctcactctagaaactgggggaaaag 780
Db 721 gtcaagcaagcaacactcaagtcacaacatgaagcgctcactctagaaactgggggaaaag 780
Qy 781 ccaaccatcgcttctcaacgaagctcctctcgaacggcagtcgggggaatcgcaaaaggat 840
Db 781 ccaaccatcgcttctcaacgaagctcctctcgaacggcagtcgggggaatcgcaaaaggat 840


```
Db 781 ccaaccatcgtcttcaacgaagctcctctcgaacggaagtcggggaaatcggcaaaaggat 840
QY 841 ttctcaaaaattcggggcaaaatttgggtccccccctcctctgttctgttagtgcaatggggaagt 900
Db 841 ttctcaaaaattcggggcaaaatttgggtccccccctcctctgttctgttagtgcaatggggaagt 900
QY 901 ttagcggagaaattcccatggagtcgcgtcatggctcatttggaggtcgtcagagatggctt 960
Db 901 ttagcggagaaattcccatggagtcgcgtcatggctcatttggaggtcgtcagagatggctt 960
QY 961 ggcagaaacccattggaacccaagagacgcgatggtcccttcgtcgaacagtcccagtac 1020
Db 961 ggcagaaacccattggaacccaagagacgcgatggtcccttcgtcgaacagtcccagtac 1020
QY 1021 gacagagctctgggttaacattgacgttggaaggaatccgcgcagctcctcactggcgctt 1080
Db 1021 gacagagctctgggttaacattgacgttggaaggaatccgcgcagctcctcactggcgctt 1080
QY 1081 ggtagaaaggcgacaaaggattcgattggaacggacgatatttgtcaatcccaaaacca 1140
Db 1081 ggtagaaaggcgacaaaggattcgattggaacggacgatatttgtcaatcccaaaacca 1140
QY 1141 ggcagcaaaatttgggttgaggagatcttggccccctctgttccattaaagacgttcaag 1200
Db 1141 ggcagcaaaatttgggttgaggagatcttggccccctctgttccattaaagacgttcaag 1200
QY 1201 acggaagaagagccattgagattgccaatgacacgacttatgggctagcctcgttcatt 1260
Db 1201 acggaagaagagccattgagattgccaatgacacgacttatgggctagcctcgttcatt 1260
QY 1261 tataccaaatctctcaacaggggtctcctcgtgtctcgtcgtcgtcgtcgtcgtcgtcgtc 1320
Db 1261 tataccaaatctctcaacaggggtctcctcgtgtctcgtcgtcgtcgtcgtcgtcgtcgtc 1320
QY 1321 tcatcaacttccctttatcccgagacacaaactcctgttggcgccatgaacaaactcg 1380
Db 1321 tcatcaacttccctttatcccgagacacaaactcctgttggcgccatgaacaaactcg 1380
QY 1381 ggctcaggcagagagctaggcggaagagggtcgaagggtcacttggagcccaagaccatt 1440
Db 1381 ggctcaggcagagagctaggcggaagagggtcgaagggtcacttggagcccaagaccatt 1440
QY 1441 aatatccacgtcaacatagagtga 1464
Db 1441 aatatccacgtcaacatagagtga 1464
```

```
RESULT 8
US-09-677-682B-4
; Sequence 4, Application US/09677682B
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; TITLE OF INVENTION: Detoxification
; FILE REFERENCE: 35718/204101
; CURRENT APPLICATION NUMBER: US/09/677,682B
; CURRENT FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1464
; TYPE: DNA
; ORGANISM: Exophiala spinifera
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)..(0)
; OTHER INFORMATION: aldehyde dehydrogenase, fully spliced cDNA
```

```
US-09-677-682B-4
Query Match 100.0%; Score 1464; DB 26; Length 1464;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1464; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 atgggttcttgccttgacgaatacaagagtgaacttctcaatcaacaatgaattcgtctcc 60
Db 1 atgggttcttgccttgacgaatacaagagtgaacttctcaatcaacaatgaattcgtctcc 60
QY 61 tccaaggggtccgagagattaaacgtcaccgaacccgtgggacgaatccaccgttgcact 120
Db 61 tccaaggggtccgagagattaaacgtcaccgaacccgtgggacgaatccaccgttgcact 120
QY 121 gatgttcaacgttgccaaacgcgcgcgtatgtcgacagtgacagtagccgcttcggtagggcg 180
Db 121 gatgttcaacgttgccaaacgcgcgcgtatgtcgacagtgacagtagccgcttcggtagggcg 180
QY 181 gtcaaaaaaggggcccatgggaagaagttcacaggtgcacaaacgcgcgcgtgcatgcttaag 240
Db 181 gtcaaaaaaggggcccatgggaagaagttcacaggtgcacaaacgcgcgcgtgcatgcttaag 240
QY 241 ttccgcggacctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 300
Db 241 ttccgcggacctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 300
QY 301 ggtagaccgggtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 360
Db 301 ggtagaccgggtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 360
QY 361 tactatgcagctgggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 420
Db 361 tactatgcagctgggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 420
QY 421 ccgaattggcgttacgagccgattgggggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 480
Db 421 ccgaattggcgttacgagccgattgggggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 480
QY 481 ttcttttacgtcggctgggaagatagcccccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 540
Db 481 ttcttttacgtcggctgggaagatagcccccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 540
QY 541 aaagcctcggagaaatccccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 600
Db 541 aaagcctcggagaaatccccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 600
QY 601 ggattccctcctggagtcgtgcagttcctcactcactggagcagcagtgacgggtggaagcattg 660
Db 601 ggattccctcctggagtcgtgcagttcctcactcactggagcagcagtgacgggtggaagcattg 660
QY 661 gcgtcgacatggacattgcgaagatcagcttcacaaagatctgtcggcggtggcgccgcgcgcgc 720
Db 661 gcgtcgacatggacattgcgaagatcagcttcacaaagatctgtcggcggtggcgccgcgcgcgc 720
QY 721 gtcaagcaagcaacactcaagtcacaaatgaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 780
Db 721 gtcaagcaagcaacactcaagtcacaaatgaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 780
QY 781 ccaaccatcgttcttcaacgaagtcctcctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 840
Db 781 ccaaccatcgttcttcaacgaagtcctcctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 840
QY 841 ttctcaaaaattcggggcaaaatttgggtccccccctcctcgttctgttagtgcaatggggaagt 900
Db 841 ttctcaaaaattcggggcaaaatttgggtccccccctcctcgttctgttagtgcaatggggaagt 900
QY 901 tttagcggagaaattcccatggagtcgcgtcatggctcatttggaggtcgtcagagatggctt 960
Db 901 tttagcggagaaattcccatggagtcgcgtcatggctcatttggaggtcgtcagagatggctt 960
QY 961 ggcagaaacccattggaacccaagagacgcgatggtcccttcgtcgaacagtcccagtac 1020
Db 961 ggcagaaacccattggaacccaagagacgcgatggtcccttcgtcgaacagtcccagtac 1020
```

Db 961 ggccagaaccattggaacccaagagagcgatggtcccttcgtgcacaaagtcccaagtac 1020
QY 1021 gacagagcttgggttaacattgacgttggaaggatataccgcgaagctcccaactgaggtt 1080
Db 1021 gacagagcttgggttaacattgacgttggaaggatataccgcgaagctcccaactgaggtt 1080
QY 1081 ggtagaaggcgacaaaggattcgcgattgaaaccgacgatatattgtcaatcccaaacca 1140
Db 1081 ggtagaaggcgacaaaggattcgcgattgaaaccgacgatatattgtcaatcccaaacca 1140
QY 1141 ggcagcaaaatttggttgaggagatcttttgccccctgttgcatttaagaacgttcaag 1200
Db 1141 ggcagcaaaatttggttgaggagatcttttgccccctgttgcatttaagaacgttcaag 1200
QY 1201 acggaagaaggcccaattgagattgccaatgacacgacttatgggctagcctcggtcatt 1260
Db 1201 acggaagaaggcccaattgagattgccaatgacacgacttatgggctagcctcggtcatt 1260
QY 1261 tatacaaaatctctcaacaggggtctccgtgtctcgtcggtcggtcggtcggtcggtc 1320
Db 1261 tatacaaaatctctcaacaggggtctccgtgtctcgtcggtcggtcggtcggtcggtc 1320
QY 1321 tcatcaacttccccctttatcccccgagacacaaactccgtttggcggtcatgaaacaatcg 1380
Db 1321 tcatcaacttccccctttatcccccgagacacaaactccgtttggcggtcatgaaacaatcg 1380
QY 1381 ggctcaggcagagagctaggcggaagggtcctcaaggcgctacttggagcccaagaccatt 1440
Db 1381 ggctcaggcagagagctaggcggaagggtcctcaaggcgctacttggagcccaagaccatt 1440
QY 1441 aatatacagctcaacatagagtga 1464
Db 1441 aatatacagctcaacatagagtga 1464

RESULT 9
US-09-882-694-4
; Sequence 4, Application US/09882694
; GENERAL INFORMATION:
; APPLICANT: Duwick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; TITLE OF INVENTION: Detoxification
; FILE REFERENCE: 35718/208255
; CURRENT APPLICATION NUMBER: US/09/882,694
; CURRENT FILING DATE: 2001-06-12
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1464
; TYPE: DNA
; ORGANISM: Exophiala spinifera
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: aldehyde dehydrogenase, fully spliced cDNA
US-09-882-694-4

Query Match 100.0%; Score 1464; DB 33; Length 1464;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1464; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 atggttcttcgcctgacgaatacaagagtgaaactcttcatcaacaatgaattcgctctcc 60
Db 1 atggttcttcgcctgacgaatacaagagtgaaactcttcatcaacaatgaattcgctctcc 60
QY 61 tccaaggggtccgagagattaaacgtctacgaaccggtggagcgaatccaccgttgccact 120

Db 61 tccaaggggtccgagagattaaacgtctacgaaccggtggacgaatccaccgttgccact 120
QY 121 gatgttcacgtggcccaacgcggccgatgtcgacagtgacgttagccgttcggtgcaggcg 180
Db 121 gatgttcacgtggcccaacgcggccgatgtcgacagtgacgttagccgttcggtgcaggcg 180
QY 181 gtcaaaaaaggcccatggaagaagtccacaggtgcacaacgcgcggcggtgcatgcttaag 240
Db 181 gtcaaaaaaggcccatggaagaagtccacaggtgcacaacgcgcggcggtgcatgcttaag 240
QY 241 ttcgcggaacctcgcgcgagagaacgcgcgagaagtcgctcgtctgagtcgctgcccacc 300
Db 241 ttcgcggaacctcgcgcgagagaacgcgcgagaagtcgctcgtctgagtcgctgcccacc 300
QY 301 ggtagaccggtgtcgtatgatactcatttcgacattccaaacatggtctccgtgttcgc 360
Db 301 ggtagaccggtgtcgtatgatactcatttcgacattccaaacatggtctccgtgttcgc 360
QY 361 tactatgcaggtggccgcacaagatgcgcgggaagaccttcccaggacaaacgcgaag 420
Db 361 tactatgcaggtggccgcacaagatgcgcgggaagaccttcccaggacaaacgcgaag 420
QY 421 ccgaattggcgttacgagccgatgggggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 480
Db 421 ccgaattggcgttacgagccgatgggggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 480
QY 481 ttctttacgtcggctggaagatagccccgcctcgcgcggcggtcgtctcttcatcttc 540
Db 481 ttctttacgtcggctggaagatagccccgcctcgcgcggcggtcgtctcttcatcttc 540
QY 541 aaagcctcggagaaaatccccgcctggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 600
Db 541 aaagcctcggagaaaatccccgcctggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 600
QY 601 ggattccctcctggtgagtcgtgcagttcctcactggagcacgagtgacgggtgaaagcattg 660
Db 601 ggattccctcctggtgagtcgtgcagttcctcactggagcacgagtgacgggtgaaagcattg 660
QY 661 gcgtcgcacatggacattgcgaagatcagcttcacaagatctgtcggcggtgcccgcgcc 720
Db 661 gcgtcgcacatggacattgcgaagatcagcttcacaagatctgtcggcggtgcccgcgcc 720
QY 721 gtcaagcaagcaacactcaagtcacaacatgaagcgctcactctagaactgggggaaag 780
Db 721 gtcaagcaagcaacactcaagtcacaacatgaagcgctcactctagaactgggggaaag 780
QY 781 ccaaccatcgtcttcaacgaagctcctcgcgaacgcagtcgggggaatcgcaagagat 840
Db 781 ccaaccatcgtcttcaacgaagctcctcgcgaacgcagtcgggggaatcgcaagagat 840
QY 841 ttctcaaaatcgggcaaaatttggtgtccccctcctctgttgcgtagtgcaatggggaat 900
Db 841 ttctcaaaatcgggcaaaatttggtgtccccctcctctgttgcgtagtgcaatggggaat 900
QY 901 ttgcggagaaaattccatggagtcctcactggctcatttggaggctgcaagatggctt 960
Db 901 ttgcggagaaaattccatggagtcctcactggctcatttggaggctgcaagatggctt 960
QY 961 ggccagaaccattggaaaccaagagagcgcagtcgtcccttcgtgcgaagtcaccagtac 1020
Db 961 ggccagaaccattggaaaccaagagagcgcagtcgtcccttcgtgcgaagtcaccagtac 1020
QY 1021 gacagagtccttgggttaacattgacgttggaaggatataccgcgaagctccactggcggtt 1080
Db 1021 gacagagtccttgggttaacattgacgttggaaggatataccgcgaagctccactggcggtt 1080
QY 1081 ggtagaaggcgacaaaggattcgcgattgaaaccgacgatatatttgtcaatcccaaacca 1140
Db 1081 ggtagaaggcgacaaaggattcgcgattgaaaccgacgatatatttgtcaatcccaaacca 1140
QY 1141 ggcagcaaaatttggttgaggagatcttttgccccctgttgcatttaagaacgttcaag 1200

```
Db 1141 ggcagcaaaatttggttgaggagatctttggcccgctcttgttcattaaagacggttcaag 1200
QY 1201 acggaagaagagggccattgagattgccaatgacagacttatgggctagcctcggtcatt 1260
Db 1201 acggaagaagagggccattgagattgccaatgacagacttatgggctagcctcggtcatt 1260
QY 1261 tataccaaatctctcaacaggggtctccggtgtctcggtcggtcggtcggtcggtcggtc 1320
Db 1261 tataccaaatctctcaacaggggtctccggtgtctcggtcggtcggtcggtcggtcggtc 1320
QY 1321 tcgatcaactcccttttatcccgagacacaaactccgttttgccgcatgaaacaatcg 1380
Db 1321 tcgatcaactcccttttatcccgagacacaaactccgttttgccgcatgaaacaatcg 1380
QY 1381 ggctcaggcagagagctaggcgaagaagggtcctcaaggcgtactttggagcccaagaccatt 1440
Db 1381 ggctcaggcagagagctaggcgaagaagggtcctcaaggcgtactttggagcccaagaccatt 1440
QY 1441 aatatccacgtcaacatagagtga 1464
Db 1441 aatatccacgtcaacatagagtga 1464

RESULT 10
US-09-882-694A-4
; Sequence 4, Application US/09882694A
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; TITLE OF INVENTION: Detoxification
; FILE REFERENCE: 5718-111
; CURRENT APPLICATION NUMBER: US/09/882.694A
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1464
; TYPE: DNA
; ORGANISM: Exophiala spinifera
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: aldehyde dehydrogenase, fully spliced cDNA
US-09-882-694A-4

Query Match 100.0%; Score 1464; DB 33; Length 1464;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1464; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggtttcttcgctgacgaatacaagagtgaaactcttcacacaaatgaattcgtctcc 60
Db 1 atggtttcttcgctgacgaatacaagagtgaaactcttcacacaaatgaattcgtctcc 60
QY 61 tccaaggggtccgagagattaaacgtcctcaacgaacccgtggagcaatccacgttgccact 120
Db 61 tccaaggggtccgagagattaaacgtcctcaacgaacccgtggagcaatccacgttgccact 120
QY 121 gatgttcacgtggccaaacggtcgatgtcgacagtgcagtagccgcttcggtgcaggcg 180
Db 121 gatgttcacgtggccaaacggtcgatgtcgacagtgcagtagccgcttcggtgcaggcg 180
QY 181 gtcaaaaagggtcccatggaagaagttcacaggtgcacaacggtcggtgcattgaag 240
Db 181 gtcaaaaagggtcccatggaagaagttcacaggtgcacaacggtcggtgcattgaag 240
QY 241 ttcgaggacctcgcgcgagaagaacgcccgaagaagctcgtcgtctgtgagtcgctgccacc 300
```

```
Db 241 ttcgaggacctcgcgcgagaagaagcccgagagctcgtcgtctggagtcgctgccacc 300
QY 301 ggtagaccggtgtcgatgatcactcatttcgacattccaaacatggtctcgtgttccgc 360
Db 301 ggtagaccggtgtcgatgatcactcatttcgacattccaaacatggtctcgtgttccgc 360
QY 361 tactatgcaggtggccgacaaagatcgccgaaagacacttcccggaggacacggcgaag 420
Db 361 tactatgcaggtggccgacaaagatcgccgaaagacacttcccggaggacacggcgaag 420
QY 421 ccgaattggcgttacgagccgagatgggggtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 480
Db 421 ccgaattggcgttacgagccgagatgggggtgtgtgtgtgtgtgtgtgtgtgtgtgt 480
QY 481 ttctttacgtggctggaagatagcccccccgctcgcgcgcgcgcgcgcgcgcgcgcgcgc 540
Db 481 ttctttacgtggctggaagatagcccccccgctcgcgcgcgcgcgcgcgcgcgcgcgcgc 540
QY 541 aaagcctcggagaaaatcccccgctggggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 600
Db 541 aaagcctcggagaaaatcccccgctggggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 600
QY 601 ggtattccctccttgagtcgtgcagttcctcactgagcagcagcagcagcagcagcagcagc 660
Db 601 ggtattccctccttgagtcgtgcagttcctcactgagcagcagcagcagcagcagcagcagc 660
QY 661 gcgtcgacatggacattgcgaagatcagcttcaacagatctgtcggcgtgtgtgtgtgtgtgt 720
Db 661 gcgtcgacatggacattgcgaagatcagcttcaacagatctgtcggcgtgtgtgtgtgtgtgt 720
QY 721 gtcaagcaagcaacactcaagttcaacatgaagcgcgtcactctagaacatgggggaaaaag 780
Db 721 gtcaagcaagcaacactcaagttcaacatgaagcgcgtcactctagaacatgggggaaaaag 780
QY 781 ccaaccatcgtcttcaacgaagtcctcctcgcgaacggtcagcagcagcagcagcagcagcagc 840
Db 781 ccaaccatcgtcttcaacgaagtcctcctcgcgaacggtcagcagcagcagcagcagcagcagc 840
QY 841 ttctcaaaattcgggcaaaatttgggtccccccctcctcgttctgtagtgcgaatggggaaat 900
Db 841 ttctcaaaattcgggcaaaatttgggtccccccctcctcgttctgtagtgcgaatggggaaat 900
QY 901 ttacggagaaaattccatggagtcctcgtcagcagcagcagcagcagcagcagcagcagcagc 960
Db 901 ttacggagaaaattccatggagtcctcgtcagcagcagcagcagcagcagcagcagcagcagc 960
QY 961 ggcagaaacccattggaaccccaagagacgcatggtccctcgtcgcgaacagtcaccagtac 1020
Db 961 ggcagaaacccattggaaccccaagagacgcatggtccctcgtcgcgaacagtcaccagtac 1020
QY 1021 gacagagtccttggttaacattgacgttggcaaggatataccgcgcagcgtcctcactggcgtt 1080
Db 1021 gacagagtccttggttaacattgacgttggcaaggatataccgcgcagcgtcctcactggcgtt 1080
QY 1081 ggtagaaaaggcgcgaaggttgcgcagattgaaacgcagcagatatttgcattcccaaacca 1140
Db 1081 ggtagaaaaggcgcgaaggttgcgcagattgaaacgcagcagatatttgcattcccaaacca 1140
QY 1141 ggcagcaaaaatttgggtttgagagatcttttggcccgctcttgcatttaagacggttcaag 1200
Db 1141 ggcagcaaaaatttgggtttgagagatcttttggcccgctcttgcatttaagacggttcaag 1200
QY 1201 acggaagaagagggccattgagattgccaatgacagacttatgggctagcctcggtcatt 1260
Db 1201 acggaagaagagggccattgagattgccaatgacagacttatgggctagcctcggtcatt 1260
QY 1261 tataccaaatctctcaacaggggtctccggtgtctcgtcgcgcgcgcgcgcgcgcgcgcgcgc 1320
Db 1261 tataccaaatctctcaacaggggtctccggtgtctcgtcgcgcgcgcgcgcgcgcgcgcgcgc 1320
QY 1321 tcgatcaacttccccctttatcccccgagacacaaactccgttggcggcatgaaacaatcg 1380
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2002, 05:54:18 ; Search time 151.01 Seconds
(without alignments)
2381.349 Million cell updates/sec

Title: US-09-882-694A-4
Perfect score: 1464
Sequence: 1 atggtctcttcgcctgacga.....tccacgtcaacatagatga 1464

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 383533 seqs, 122816752 residues

Word size : 0

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA:*

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
- 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
- 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
- 5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	18	1.2	1125	4	US-09-516-914-12
2	18	1.2	1262	2	US-08-989-386-4
3	18	1.2	1839	1	US-08-442-248-3
4	18	1.2	1839	1	US-08-440-815-3
5	18	1.2	1839	4	US-08-486-449-3
6	18	1.2	14272	4	US-09-516-914-23
7	17	1.2	111	4	US-09-127-680-11
8	17	1.2	340	3	US-08-441-971-10
9	17	1.2	340	3	US-08-441-971-11
10	17	1.2	340	4	US-08-221-653-10
11	17	1.2	340	4	US-08-221-653-11
12	17	1.2	340	4	US-08-442-144A-10
13	17	1.2	340	4	US-08-442-144A-11
14	17	1.2	340	4	US-08-441-970-10
15	17	1.2	340	4	US-08-441-970-11
16	17	1.2	1079	1	US-08-471-570-13
c 17	17	1.2	1950	2	US-08-377-440A-2
c 18	17	1.2	1950	4	US-09-440-530-2
19	17	1.2	1954	1	US-08-471-570-5
20	17	1.2	2181	4	US-09-158-767-10
21	17	1.2	2676	1	US-08-471-570-7
22	17	1.2	2991	1	US-08-324-977-49
23	17	1.2	2991	2	US-08-384-616-49
24	17	1.2	2991	2	US-08-904-686A-49
25	17	1.2	2991	4	US-09-315-850-49
26	17	1.2	3416	2	US-08-451-822A-15
27	17	1.2	3416	4	US-08-323-430-15

28	17	1.2	7001	1	US-08-258-261B-1	Sequence 1, Appli
29	17	1.2	7001	1	US-08-456-837-1	Sequence 1, Appli
30	17	1.2	7001	1	US-08-457-342-1	Sequence 1, Appli
31	17	1.2	7001	1	US-08-457-646A-1	Sequence 1, Appli
32	17	1.2	7001	1	US-08-458-076A-1	Sequence 1, Appli
33	17	1.2	7001	1	US-08-457-335A-1	Sequence 1, Appli
34	17	1.2	7001	1	US-08-761-258-6	Sequence 6, Appli
35	17	1.2	7001	1	US-08-729-214-1	Sequence 1, Appli
36	17	1.2	7001	2	US-08-977-306-6	Sequence 6, Appli
37	17	1.2	7001	3	US-09-028-934-1	Sequence 1, Appli
38	17	1.2	7863	1	US-08-324-977-35	Sequence 35, Appli
39	17	1.2	7863	2	US-08-384-616-35	Sequence 35, Appli
40	17	1.2	7863	2	US-08-904-686A-35	Sequence 35, Appli
41	17	1.2	7863	4	US-09-315-850-35	Sequence 35, Appli
42	17	1.2	7917	1	US-08-324-977-31	Sequence 31, Appli
43	17	1.2	7917	2	US-08-384-616-31	Sequence 31, Appli
44	17	1.2	7917	2	US-08-904-686A-31	Sequence 31, Appli
45	17	1.2	7917	4	US-09-315-850-31	Sequence 31, Appli

ALIGNMENTS

RESULT 1
US-09-516-914-12/C
; Sequence 12, Application US/09516914
; Patent No. 6333401
; GENERAL INFORMATION:
; APPLICANT: Breinig, Sabine
; APPLICANT: Fuchs, Georg
; TITLE OF INVENTION: Phenol-induced Proteins of Thauera aromatica
; FILE REFERENCE: BC1006 US NA
; CURRENT APPLICATION NUMBER: US/09/516.914
; CURRENT FILING DATE: 2000-03-01
; EARLIER APPLICATION NUMBER: 60/122,952
; EARLIER FILING DATE: 1999-03-05
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 12
; LENGTH: 1125
; TYPE: DNA
; ORGANISM: Thauera aromatica
US-09-516-914-12

Query Match 1.2%; Score 18; DB 4; Length 1125;
Best Local Similarity 100.0%; Pred. NO. 8.5;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 701 ctgtcgcggtggtgcgcgcg 718
Db 50 CTGTGGCGGTGCGCG 33

RESULT 2
US-08-989-386-4
; Sequence 4, Application US/08989386
; Patent No. 5989860
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: HUMAN ISOMERASE HOMOLOGS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
FILING DATE: US/08/989,386
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0443 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1262 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: TMLRDT01
CLONE: 292808
US-08-989-386-4

Query Match 1.2%; Score 18; DB 2; Length 1262;
Best Local Similarity 100.0%; Pred. No. 8.5;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 767 aactggggggaagccaa 784
|||||
Db 956 AACTGGGGGAAGCCAA 973

RESULT 3
US-08-442-248-3
Sequence 3, Application US/08442248
Patent No. 5759863
GENERAL INFORMATION:
APPLICANT: Caras, Ingrid W.
APPLICANT: Winslow, John W.
TITLE OF INVENTION: AL-1 Neurotrophic Factor
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/442,248
FILING DATE: 15-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/330128
FILING DATE: 27-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: Torchia, Timothy E.
REGISTRATION NUMBER: 36,700
REFERENCE/DOCKET NUMBER: 920C4

TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-8674
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1839 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-442-248-3

Query Match 1.2%; Score 18; DB 1; Length 1839;
Best Local Similarity 100.0%; Pred. No. 8.6;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 506 ccccgccctgcgcgcg 523
|||||
Db 6 CCCCCCCTCGCGCGG 23

RESULT 4
US-08-440-815-3
Sequence 3, Application US/08440815
Patent No. 5798448
GENERAL INFORMATION:
APPLICANT: Caras, Ingrid W.
APPLICANT: Winslow, John W.
TITLE OF INVENTION: AL-1 Neurotrophic Factor
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/440,815
FILING DATE: 15-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/330128
FILING DATE: 27-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: Torchia, Timothy E.
REGISTRATION NUMBER: 36,700
REFERENCE/DOCKET NUMBER: 920C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-8674
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1839 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-440-815-3

Query Match 1.2%; Score 18; DB 1; Length 1839;
Best Local Similarity 100.0%; Pred. No. 8.6;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 506 ccccgccctgcgcgcg 523
|||||

ZIP: 02210
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch
COMPUTER: IBM compatible
OPERATING SYSTEM: MS-DOS Version 3.3
SOFTWARE: WordPerfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/441,971
FILING DATE: 16-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/221,653
FILING DATE:
APPLICATION NUMBER: US/07/881,528
FILING DATE:
APPLICATION NUMBER: 07/697,326
FILING DATE: 8 May 1991
ATTORNEY/AGENT INFORMATION:
NAME: Janluk, Anthony J.
REGISTRATION NUMBER: 29,809
REFERENCE/DOCKET NUMBER: C0772/7000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 720-3500
TELEFAX: (617) 720-2441
TELEX: EZEKIEL
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 340 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: ns5gh6
US-08-441-971-10

Query Match 1.2%; Score 17; DB 3; Length 340;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 975 ggaacccaagaggagcgc 991
|||||
DB 308 GGAACCCAAGAGGACGC 324

RESULT 9
US-08-441-971-11
Sequence 11, Application US/08441971
Patent No. 6071693
GENERAL INFORMATION:
APPLICANT: Tai-An Cha
TITLE OF INVENTION: HCV GENOMIC SEQUENCES FOR
DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 147
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
STREET: 600 Atlantic Avenue
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02210
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch
COMPUTER: IBM compatible
OPERATING SYSTEM: MS-DOS Version 3.3
SOFTWARE: WordPerfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/441,971
FILING DATE: 16-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/221,653

FILING DATE:
APPLICATION NUMBER: US/07/881,528
FILING DATE:
APPLICATION NUMBER: 07/697,326
FILING DATE: 8 May 1991
ATTORNEY/AGENT INFORMATION:
NAME: Janluk, Anthony J.
REGISTRATION NUMBER: 29,809
REFERENCE/DOCKET NUMBER: C0772/7000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 720-3500
TELEFAX: (617) 720-2441
TELEX: EZEKIEL
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 340 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: ns5spl
US-08-441-971-11

Query Match 1.2%; Score 17; DB 3; Length 340;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 975 ggaacccaagaggagcgc 991
|||||
DB 308 GGAACCCAAGAGGACGC 324

RESULT 10
US-08-221-653-10
Sequence 10, Application US/08221653
Patent No. 6190864
GENERAL INFORMATION:
APPLICANT: Tai-An Cha
TITLE OF INVENTION: HCV GENOMIC SEQUENCES FOR
DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 147
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
STREET: 600 Atlantic Avenue
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02210
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch
COMPUTER: IBM compatible
OPERATING SYSTEM: MS-DOS version 3.3
SOFTWARE: WordPerfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/221,653
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/881,528
FILING DATE:
APPLICATION NUMBER: 07/697,326
FILING DATE: 8 May 1991
ATTORNEY/AGENT INFORMATION:
NAME: Janluk, Anthony J.
REGISTRATION NUMBER: 29,809
REFERENCE/DOCKET NUMBER: C0772/7000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 720-3500
TELEFAX: (617) 720-2441
TELEX: EZEKIEL
INFORMATION FOR SEQ ID NO: 10:


```

; APPLICANT: Eileen Beall
; APPLICANT: Bruce Irvine
; APPLICANT: Janice Kolberg
; APPLICANT: Michael S. Urdea
; TITLE OF INVENTION: HCV GENOMIC SEQUENCES FOR
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: California
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 Inch
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows NT
; SOFTWARE: Microsoft Word 97
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/442,144A
; FILING DATE: MAY 16, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/221,653
; FILING DATE: APRIL 1, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Doreen Yatko Trujillo
; REGISTRATION NUMBER: 35,719
; REFERENCE/DOCKET NUMBER: CHIR-0121
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; TELEX:
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 340 Nucleotides
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; MOLECULE TYPE: DNA
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: ns5spl
; US-08-442-144A-11

Query Match 1.2%; Score 17; DB 4; Length 340;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 975 ggaacccaagaggacgc 991
   |||||
Db 308 GGAACCCAAGAGGACGC 324

RESULT 14
US-08-441-970-10
; Sequence 10, Application US/08441970
; Patent No. 6297370
; GENERAL INFORMATION:
; APPLICANT: Tai-An Cha
; TITLE OF INVENTION: HCV GENOMIC SEQUENCES FOR
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS Version 3.3
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/441,970
; FILING DATE: 16-MAY-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/881,528
; FILING DATE: 08-MAY-1992
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 8 May 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Janiuk, Anthony J.

```

```

; COMPUTER: IBM compatible
; OPERATING SYSTEM: MS-DOS Version 3.3
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/441,970
; FILING DATE: 16-MAY-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/881,528
; FILING DATE: 08-MAY-1992
; APPLICATION NUMBER: 07/697,326
; FILING DATE: 8 May 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Janiuk, Anthony J.
; REGISTRATION NUMBER: 29,809
; REFERENCE/DOCKET NUMBER: C0772/7000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 720-3500
; TELEFAX: (617) 720-2441
; TELEX: EZEKIEL
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 340 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: ns5gh6
; US-08-441-970-10

Query Match 1.2%; Score 17; DB 4; Length 340;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 975 ggaacccaagaggacgc 991
   |||||
Db 308 GGAACCCAAGAGGACGC 324

RESULT 15
US-08-441-970-11
; Sequence 11, Application US/08441970
; Patent No. 6297370
; GENERAL INFORMATION:
; APPLICANT: Tai-An Cha
; TITLE OF INVENTION: HCV GENOMIC SEQUENCES FOR
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch
; COMPUTER: IBM compatible
; OPERATING SYSTEM: MS-DOS Version 3.3
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/441,970
; FILING DATE: 16-MAY-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/881,528
; FILING DATE: 08-MAY-1992
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 8 May 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Janiuk, Anthony J.

```

```

; REGISTRATION NUMBER: 29,809
; REFERENCE/DOCKET NUMBER: C0772/7000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 720-3500
; TELEFAX: (617) 720-2441
; TELEX: EZEKIEL
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 340 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: ns5spl
US-08-441-970-11

```

```

Query Match      1.2%; Score 17; DB 4; Length 340;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 975 ggaacccaagagacgc 991
   |||||
Db 308 GGAACCCAAGAGGACGC 324

```

Search completed: April 27, 2002, 05:54:29
Job time: 15356 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2002, 05:28:04 ; Search time 705 Seconds
(without alignments)
3565.337 Million cell updates/sec

Title: US-09-882-694A-4
Perfect score: 1464
Sequence: 1 atggttctttgcgcctgacga.....tccacgtcaacatagagtga 1464

Scoring table: OLIGO_NUC
Gapop 60.0., Gapext 60.0

Searched: 1736436 seqs, 858457221 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:*
23: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT:*
24: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1464	100.0	1464	22 AAD02692	Exophiala spinifer
2	20	1.4	1980	23 ABL10879	Drosophila melanog
C 3	20	1.4	3278	23 ABL24684	Drosophila melanog
C 4	20	1.4	6153	23 ABL10878	Drosophila melanog
5	19	1.3	286	21 AAC03035	Human secreted pro
6	19	1.3	578	22 AAF68055	Corynebacterium gl
C 7	19	1.3	600	22 AAF22587	Human breast cance
8	19	1.3	828	23 ABL17815	Drosophila melanog
C 9	19	1.3	1015	22 AAF22491	Human breast cance

10	19	1.3	1887	22 AAH67550	C glutamicum codin
11	19	1.3	2010	22 AAF68054	Corynebacterium gl
12	19	1.3	3417	23 ABL17814	Drosophila melanog
13	19	1.3	6006	23 ABL17820	Drosophila melanog
14	19	1.3	6199	22 AAS40631	DNA encoding human
15	19	1.3	6199	22 AAL06519	Human reproductive
16	19	1.3	349980	22 AAH68532	C glutamicum codin
17	18	1.2	365	22 AAH57360	Human brain specif
C 18	18	1.2	544	22 AAK92455	Human cDNA 3'-end
C 19	18	1.2	843	23 ABL21765	Drosophila melanog
C 20	18	1.2	1125	21 AAA75090	Open reading frame
21	18	1.2	1146	22 AAF61289	N. magadali bacter
22	18	1.2	1262	21 AAZ32926	Human isomerase ho
23	18	1.2	1299	20 AAX19492	Human secreted pro
24	18	1.2	1410	21 AAC77073	Human ORFX ORF2628
25	18	1.2	1488	16 AAQ86275	Alternaria alterna
26	18	1.2	1532	21 AAZ46950	Rat B5 receptor en
C 27	18	1.2	1560	23 AAS77657	DNA encoding novel
28	18	1.2	1560	23 AAS88781	DNA encoding novel
C 29	18	1.2	1560	23 AAS91130	DNA encoding novel
30	18	1.2	1710	21 AAF07491	Fusarium venenatum
31	18	1.2	1839	17 AAT18897	Human AL-1 cDNA.
32	18	1.2	1938	22 AAH17691	Human cDNA sequenc
33	18	1.2	1968	23 ABL19437	Drosophila melanog
34	18	1.2	1999	23 ABL20087	Drosophila melanog
35	18	1.2	2282	22 AAK94158	Human full-length
36	18	1.2	2340	22 AAS31168	Human diagnostic a
37	18	1.2	2625	22 AAS44597	Human full-length
38	18	1.2	2644	22 AAS44769	Human contig polyn
39	18	1.2	3192	23 ABL21764	Drosophila melanog
C 40	18	1.2	4113	23 ABL20086	Drosophila melanog
C 41	18	1.2	4633	23 ABL05694	Drosophila melanog
C 42	18	1.2	4727	23 ABL19436	Drosophila melanog
C 43	18	1.2	6412	24 AAS61146	Human gene regulat
C 44	18	1.2	9729	22 AAS14089	Human FCTR3f DNA s
C 45	18	1.2	9826	22 AAS14085	Human FCTR3b DNA s

ALIGNMENTS

RESULT 1

AAD02692
ID AAD02692 standard; DNA; 1464 BP.

XX AAD02692;

DT 02-MAY-2001 (first entry)

DE Exophiala spinifera aldehyde dehydrogenase coding sequence.

XX Fumonisin; degradative enzyme; transport enzyme; flavin monooxygenase;
KW aldehyde dehydrogenase; permease; p-glycoprotein; fumonisin esterase;
KW detoxification; mycotoxin; animal feed; human feed; silage;
KW transgenic plant; transgenic animal; microbial spray; ds.

XX OS Exophiala spinifera.

XX FH Key Location/Qualifiers

FT CDS 1..1464
FT /*tag= a
FT /product= "E. spinifera aldehyde dehydrogenase"

XX WO200105980-A1.

XX PD 25-JAN-2001.

XX PF 14-JUL-1999; 99WO-US15824.

XX PR 14-JUL-1999; 99WO-US15824.

XX (PION-) PIONEER HI-BRED INT INC.

PA (CURA-) CURAGEN CORP.

AC ABL10879;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 27119.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -
PT
PT
XX
PS Claim 1; SEQ ID NO 27119; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABBS57737-ABBS72072).
CC
CC The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1980 BP; 471 A; 527 C; 540 G; 442 T; 0 other;

Query Match 1.4%; Score 20; DB 23; Length 1980;
Best Local Similarity 100.0%; Pred. No. 8.1;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1159 gaggagatctttggcccggt 1178
Db 1681 gaggagatctttggcccggt 1700
|||||

RESULT 3
ABL24684/C
ID ABL24684 standard; DNA; 3278 BP.
XX
AC ABL24684;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 25525.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.

XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -
PT
PT
XX
PS Claim 1; SEQ ID NO 25525; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABBS57737-ABBS72072).
CC
CC The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 3278 BP; 922 A; 729 C; 679 G; 948 T; 0 other;

Query Match 1.4%; Score 20; DB 23; Length 3278;
Best Local Similarity 100.0%; Pred. No. 8.1;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1159 gaggagatctttggcccggt 1178
Db 3046 GAGGAGATCTTTGGCCCCGT 3027
|||||

RESULT 4
ABL10878/C
ID ABL10878 standard; cDNA; 6153 BP.
XX
AC ABL10878;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 27116.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.

DR P-PSDB; ABB66775.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 27116; 21pp + Sequence Listing; English.
PS
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 6153 BP; 1689 A; 1322 C; 1355 G; 1787 T; 0 other;

Query Match 1.4%; Score 20; DB 23; Length 6153;
Best Local Similarity 100.0%; Pred. No. 8;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1159 gaggagatcttggcccg 1178
|||||
DB 1300 GAGGAGATCTTGGCCCCGT 1281

RESULT 5
AAC03035
ID AAC03035 standard; cDNA; 286 BP.
XX
AC AAC03035;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 3033.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-500381/45.
DR P-PSDB; AAG03029.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 1; SEQ ID 3033; 71pp + CD-ROM; English.
XX

The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. An ORF has been identified within the
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC derived from 30 different tissues. EST sequences usually correspond
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are

CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC well suited for isolating cDNA sequences derived from the 5' ends of
CC mRNAs and even in those cases where longer cDNA sequences have been
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC mRNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC gene therapy and chromosome mapping procedures. They are used to obtain
CC upstream regulatory sequences and to design expression and secretion
CC vectors.
XX
SQ Sequence 286 BP; 59 A; 89 C; 50 G; 84 T; 4 other;

Query Match 1.3%; Score 19; DB 21; Length 286;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 528 ctcttcattcttcaagcc 546
|||||
DB 116 ctcttcattcttcaagcc 134

RESULT 6
AAF68055
ID AAF68055 standard; DNA; 578 BP.
XX
AC AAF68055;
XX
DT 11-APR-2001 (first entry)
XX
DE Corynebacterium glutamicum MCT protein encoding DNA SEQ ID NO:625.
XX
KW Corynebacterium glutamicum; brevibacterium lactofermentum; MCT;
KW membrane construction and membrane transport protein; petroleum spill;
KW hydrocarbon degradation; gram positive aerobic bacterium; marker;
KW identification; microorganism; fine chemical production; transformation;
KW genome mapping; genetic engineering; ds.
XX
OS Corynebacterium glutamicum.
XX
PN WO200100805-A2.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000WO-IB00926.
XX
PR 25-JUN-1999; 99US-0141031.
PR 08-JUL-1999; 99DE-1031454.
PR 08-JUL-1999; 99DE-1031478.
PR 08-JUL-1999; 99DE-1031563.
PR 09-JUL-1999; 99DE-1032122.
PR 09-JUL-1999; 99DE-1032124.
PR 09-JUL-1999; 99DE-1032125.
PR 09-JUL-1999; 99DE-1032128.
PR 09-JUL-1999; 99DE-1032180.
PR 09-JUL-1999; 99DE-1032182.
PR 09-JUL-1999; 99DE-1032190.
PR 09-JUL-1999; 99DE-1032191.
PR 09-JUL-1999; 99DE-1032209.
PR 09-JUL-1999; 99DE-1032212.
PR 09-JUL-1999; 99DE-1032227.
PR 09-JUL-1999; 99DE-1032228.
PR 09-JUL-1999; 99DE-1032229.
PR 09-JUL-1999; 99DE-1032230.
PR 14-JUL-1999; 99DE-1032927.
PR 14-JUL-1999; 99DE-1033005.
PR 14-JUL-1999; 99DE-1033006.
PR 27-AUG-1999; 99DE-1040764.
PR 27-AUG-1999; 99DE-1040765.
PR 27-AUG-1999; 99DE-1040766.
PR 27-AUG-1999; 99DE-1040830.
PR 27-AUG-1999; 99DE-1040831.
PR 27-AUG-1999; 99DE-1040832.

PR 27-AUG-1999; 99DE-1040833.
PR 31-AUG-1999; 99DE-1041378.
PR 31-AUG-1999; 99DE-1041379.
PR 31-AUG-1999; 99DE-1041395.
PR 03-SEP-1999; 99DE-1042077.
PR 03-SEP-1999; 99DE-1042078.
PR 03-SEP-1999; 99DE-1042079.
PR 03-SEP-1999; 99DE-1042088.
XX
PA (BADI) BASF AG.
XX
PI Pompejus M, Kroeger B, Schroeder H, Zeider O, Haberhauer G;
XX
DR WPI; 2001-071486/08.
DR P-PSDB; AAB76822.
XX
PT Corynebacterium glutamicum nucleic acids encoding membrane construction
XX and membrane transport proteins or their portions, useful for typing or
PT identifying C. glutamicum or related bacteria, and as markers for
PT transformation -
XX
PS Claim 3; Page 1048-1049; 1119pp; English.
XX
CC AAF67743 to AAF68080 encode the Corynebacterium glutamicum membrane
CC construction and membrane transport (MCT) proteins given in AAB76510 to
CC AAB76847. The MCT nucleic acids and proteins are useful in the
CC identification of microorganisms which can be used to produce fine
CC chemicals, for modulating fine chemical production in C. glutamicum or
CC related bacteria (e.g. Brevibacterium lactofermentum), the typing or
CC identification of C. glutamicum or related bacteria, as reference points
CC for mapping C. glutamicum genome, and as markers for transformation.
CC AAF68082 and AAF68082 represent sequencing primers which are used in an
CC example from the present invention.
XX
SQ Sequence 578 BP; 146 A; 182 C; 136 G; 114 T; 0 other;

Query Match 1.3%; Score 19; DB 22; Length 578;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1215 cattgagattgccaatgac 1233
|||||
Db 456 cattgagattgccaatgac 474

RESULT 7
AAF22587/c
ID AAF22587 standard; cDNA; 600 BP.
XX
AC AAF22587;
XX
DT 26-MAR-2001 (first entry)
XX
DE Human breast cancer associated antigen nucleotide sequence SEQ ID NO:166.
XX
DE Human; breast cancer; gastric cancer; prostate cancer; diagnosis;
KW cancer associated antigen; cytostatic; cancer vaccine; ss.
KW
XX Homo sapiens.
OS
XX WO200073801-A2.
PN
XX
PD 07-DEC-2000.
XX
PF 26-MAY-2000; 2000WO-US14749.
XX
PR 28-MAY-1999; 99US-0136526.
PR 10-SEP-1999; 99US-0153454.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Obata Y;

XX WPI; 2001-025274/03.
DR
XX Nucleic acids encoding breast, gastric and prostate cancer associated
PT antigen precursors, useful for diagnosing and treating a condition
PT characterized by expression of an abnormal amount of a protein, e.g.
PT cancer -
XX
PS Claim 50; Page 328; 799pp; English.
XX
CC AAF22422 to AAF22626, AAF22627 to AAF22773 and AAF22774 to AAF23014
CC represent nucleotide sequences encoding human breast, gastric and
CC prostate cancer associated antigen precursors (CAAP) respectively.
CC AAB63232 to AAB63467, AAB63468 to AAB63721 and AAB63722 to AAB63970
CC represent human breast, gastric and prostate CAAP protein sequence
CC respectively. CAAPs have cytostatic activity and can be used in the
CC production of cancer vaccines. The human CAAP proteins, peptides, nucleic
CC acids or anti-CAAP antibodies are useful for diagnosing and treating a
CC condition characterised by expression of an abnormal amount of a protein,
CC e.g. cancer.
XX
SQ Sequence 600 BP; 103 A; 116 C; 113 G; 193 T; 75 other;

Query Match 1.3%; Score 19; DB 22; Length 600;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 tcaacttcccccttatccc 1343
|||||
Db 451 TCAACTTCCCCTTTATCCC 433

RESULT 8
ABL17815
ID ABL17815 standard; DNA; 828 BP.
XX
AC ABL17815;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 4918.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 4918; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX
SQ Sequence 828 BP; 161 A; 241 C; 237 G; 189 T; 0 other;

Query Match 1.3%; Score 19; DB 23; Length 828;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 172 gtgcaggcggtcaaaaagg 190
|||||
Db 778 gtgcaggcggtcaaaaagg 796
|||||

RESULT 9
AAF22491/c
ID AAF22491 standard; cDNA; 1015 BP.
XX
AC AAF22491;
XX
DT 26-MAR-2001 (first entry)
XX
DE Human breast cancer associated antigen nucleotide sequence SEQ ID NO:70.
XX
KW Human; breast cancer; gastric cancer; prostate cancer; diagnosis;
KW cancer associated antigen; cytostatic; cancer vaccine; ss.
XX
OS Homo sapiens.
XX
PN WO200073801-A2.
XX
PD 07-DEC-2000.
XX
PF 26-MAY-2000; 2000WO-US14749.
XX
PR 28-MAY-1999; 99US-0136526.
PR 10-SEP-1999; 99US-0153454.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Obata Y;
XX
DR WPI; 2001-025274/03.
XX
PT Nucleic acids encoding breast, gastric and prostate cancer associated
PT antigen precursors, useful for diagnosing and treating a condition
PT characterized by expression of an abnormal amount of a protein, e.g.
PT cancer -
XX
PS Claim 50; Page 297; 799pp; English.
XX
CC AAF22422 to AAF22626, AAF22627 to AAF22773 and AAF22774 to AAF23014
CC represent nucleotide sequences encoding human breast, gastric and
CC prostate cancer associated antigen precursors (CAAP) respectively.
CC AAB63232 to AAB63467, AAB63468 to AAB63721 and AAB63722 to AAB63970
CC represent human breast, gastric and prostate CAAP protein sequence
CC respectively. CAAPs have cytostatic activity and can be used in the
CC production of cancer vaccines. The human CAAP proteins, peptides, nucleic
CC acids or anti-CAAP antibodies are useful for diagnosing and treating a
CC condition characterised by expression of an abnormal amount of a protein,
CC e.g. cancer.
XX
SQ Sequence 1015 BP; 192 A; 179 C; 186 G; 340 T; 118 other;

Query Match 1.3%; Score 19; DB 22; Length 1015;
Best Local Similarity 100.0%; Pred. No. 26;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1325 tcaacttccccctttatccc 1343
|||||
Db 451 TCAACTTCCCCCTTTATCCC 433
|||||

RESULT 10
AAH67550
ID AAH67550 standard; DNA; 1887 BP.
XX
AC AAH67550;
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum coding sequence fragment SEQ ID NO: 2585.
XX
KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis; ds.
XX
OS Corynebacterium glutamicum.
XX
PN EPI108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-0127688.
XX
PR 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
DR WPI; 2001-376931/40.
DR P-PSDB; AAG92331.
XX
PT Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT -
XX
PS Claim 8; SEQ ID NO: 2585; 246pp + Sequence Listing; English.
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a nucleic acid described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 1887 BP; 338 A; 576 C; 498 G; 475 T; 0 other;

Query Match 1.3%; Score 19; DB 22; Length 1887;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1215 cattgagattgccaatgac 1233
|||||
Db 1788 cattgagattgccaatgac 1806
|||||

RESULT 11
AAF68054
ID AAF68054 standard; DNA; 2010 BP.
XX
AC AAF68054;
XX
DT 11-APR-2001 (first entry)
XX
DE Corynebacterium glutamicum MCT protein encoding DNA SEQ ID NO:623.
XX
DE Corynebacterium glutamicum; brevibacterium lactofermentum; MCT;
KW membrane construction and membrane transport protein; petroleum spill;
KW hydrocarbon degradation; gram positive aerobic bacterium; marker;
KW identification; microorganism; fine chemical production; transformation;
KW genome mapping; genetic engineering; ds.
XX
OS Corynebacterium glutamicum.
XX
PN WO200100805-A2.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000WO-IB00926.
XX
PR 25-JUN-1999; 99US-0141031.
PR 08-JUL-1999; 99DE-1031454.
PR 08-JUL-1999; 99DE-1031478.
PR 08-JUL-1999; 99DE-1031563.
PR 09-JUL-1999; 99DE-1032122.
PR 09-JUL-1999; 99DE-1032124.
PR 09-JUL-1999; 99DE-1032125.
PR 09-JUL-1999; 99DE-1032128.
PR 09-JUL-1999; 99DE-1032180.
PR 09-JUL-1999; 99DE-1032182.
PR 09-JUL-1999; 99DE-1032190.
PR 09-JUL-1999; 99DE-1032191.
PR 09-JUL-1999; 99DE-1032209.
PR 09-JUL-1999; 99DE-1032212.
PR 09-JUL-1999; 99DE-1032227.
PR 09-JUL-1999; 99DE-1032228.
PR 09-JUL-1999; 99DE-1032230.
PR 09-JUL-1999; 99DE-1032230.
PR 14-JUL-1999; 99DE-1032927.
PR 14-JUL-1999; 99DE-1033005.
PR 14-JUL-1999; 99DE-1033006.
PR 27-AUG-1999; 99DE-1040764.
PR 27-AUG-1999; 99DE-1040765.
PR 27-AUG-1999; 99DE-1040766.
PR 27-AUG-1999; 99DE-1040830.
PR 27-AUG-1999; 99DE-1040831.
PR 27-AUG-1999; 99DE-1040832.
PR 27-AUG-1999; 99DE-1040833.
PR 31-AUG-1999; 99DE-1041378.
PR 31-AUG-1999; 99DE-1041379.
PR 31-AUG-1999; 99DE-1041395.
PR 03-SEP-1999; 99DE-1042077.
PR 03-SEP-1999; 99DE-1042078.
PR 03-SEP-1999; 99DE-1042079.
PR 03-SEP-1999; 99DE-1042088.
XX
PA (BADI) BASF AG.
XX
PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
XX
DR WPI; 2001-071486/08.
DR P-PSDB; AAB76821.
XX
PT Corynebacterium glutamicum nucleic acids encoding membrane construction
PT and membrane transport proteins or their portions, useful for typing or
PT identifying C. glutamicum or related bacteria, and as markers for
PT transformation -
XX
PS Claim 3; Page 1043-1046; 1119pp; English.

XX AAF67743 to AAF68080 encode the Corynebacterium glutamicum membrane
CC construction and membrane transport (MCT) proteins given in AAB76510 to
CC AAB76847. The MCT nucleic acids and proteins are useful in the
CC identification of microorganisms which can be used to produce fine
CC chemicals, for modulating fine chemical production in C. glutamicum or
CC related bacteria (e.g. Brevibacterium lactofermentum), the typing or
CC identification of C. glutamicum or related bacteria, as reference points
CC for mapping C. glutamicum genome, and as markers for transformation.
CC AAF68082 and AAF68082 represent sequencing primers which are used in an
CC example from the present invention.
XX
SQ Sequence 2010 BP; 389 A; 607 C; 522 G; 492 T; 0 other;

Query Match 1.3%; Score 19; DB 22; Length 2010;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1215 cattgagattgccaatgac 1233
 |||||
Db 1888 cattgagattgccaatgac 1906

RESULT 12
ABLL7814
ID ABL17814 standard; DNA; 3417 BP.
XX
AC ABL17814;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 4915.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 4915; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 3417 BP; 787 A; 948 C; 920 G; 762 T; 0 other;

PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235844.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.

PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
PI
XX WPI; 2001-451929/48.
DR
XX
XX Isolated polypeptide for treating, preventing and/or prognosing
PT disorders related to the reproductive system including prostate cancer
PT and also for testing and detection e.g. diagnosis -
XX
PS Disclosure; SEQ ID No 783; 546pp; English.
XX
CC The invention relates to novel isolated human prostate cancer antigen
CC polynucleotides (I) and polypeptides (II). (I) and (II) are useful for
CC preventing, treating or ameliorating a medical condition when
CC administered. (I), (II) and the antibody to (II) are useful for treating,
CC preventing and/or prognosing disorders related to the reproductive
CC system including prostate cancers; urinary disorders e.g. chronic
CC nephritis; and blood-related disorders e.g. thrombosis. (II) can be used
CC for testing and detection e.g. as a chromosomal marker and in forensics.
CC (I) and the anti-(II) antibody can be used in testing and detection in
CC immunoassays. AAS40061-AAS40775 represent the human prostate cancer
CC antigen coding sequences, and related PCR primers and sequences of the
CC invention. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at: ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 6199 BP; 1516 A; 1686 C; 1432 G; 1565 T; 0 other;

Query Match 1.3%; Score 19; DB 22; Length 6199;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1211 aggccattgagattgcca 1229
Db 1648 aggccattgagattgcca 1666
|||||

RESULT 15
AAL06519
ID AAL06519 standard; DNA; 6199 BP.
XX
AC AAL06519;
XX
DT 21-NOV-2001 (first entry)
XX
DE Human reproductive system related antigen DNA SEQ ID NO: 9207.
XX
KW Human; reproductive system related antigen; reproductive system disorder;
KW cancer; gene therapy; ds.
XX
OS Homo sapiens.
XX
PN WO200155320-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01339.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.

PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-465570/50.
XX
XX Isolated nucleic acid molecule encoding a reproductive system antigen
PT is used in preventing, treating or ameliorating a medical condition -
XX
PS Disclosure; SEQ ID NO 9207; 1297pp + Sequence Listing; English.
XX

CC The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a genomic sequence encoding a
CC protein of the invention.

XX
SQ Sequence 6199 BP; 1516 A; 1686 C; 1432 G; 1565 T; 0 other;

Query Match 1.3%; Score 19; DB 22; Length 6199;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1211 aggccattgagattgccaa 1229
|
Db 1648 aggccattgagattgccaa 1666

Search completed: April 27, 2002, 05:28:23
Job time: 13786 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic ~ nucleic search, using sw model

Run on: April 27, 2002, 05:23:55 ; Search time 7820.58 Seconds
(without alignments)
3917.414 Million cell updates/sec

Title: US-09-882-694A-4
Perfect score: 1464
Sequence: 1 atggttcttttcgctgacga.....tccacgtcaacatagagtga 1464

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0
Searched: 1797656 seqs, 10463268293 residues

Word size : 0
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

1	1464	100.0	1464	6	AX076845	AX076845 Sequence
2	24	1.6	4158	8	ABY17825	Y17825 Agaricus bi
3	22	1.5	157201	8	AP003048	AP003048 Oryza sat
4	22	1.5	260050	1	RME603643	AL603643 Rhizobium
5	21	1.4	3897	9	HSU62769	U62769 Human oxyto
6	21	1.4	4226	9	HSAL131024	AJ131024 Homo sapi
7	21	1.4	10498	1	AB008831	AB008831 Acinetoba
8	21	1.4	102910	9	AL356308	AL356308 Human DNA
c 9	21	1.4	131805	9	AC008850	AC008850 Homo sapi
c 10	21	1.4	138904	2	AP003928	AP003928 Oryza sat
11	21	1.4	147344	2	AC008907	AC008907 Homo sapi
12	21	1.4	191415	2	AL627302	AL627302 Mus muscu
c 13	21	1.4	251664	2	AC008908	AC008908 Homo sapi
14	20	1.4	481	14	FVNS5GAI	L48981 Sepik virus
15	20	1.4	2001	1	AB009654	AB009654 Alteromon
c 16	20	1.4	11553	1	AE005775	AE005775 Caulobact
17	20	1.4	20701	1	AE006958	AE006958 Mycobacte
18	20	1.4	29550	1	MTV039	AL021942 Mycobacte
c 19	20	1.4	31087	2	AC014839	AC014839 Drosophil
c 20	20	1.4	48986	2	AC105976	AC105976 Mus muscu
c 21	20	1.4	58005	2	AC109128	AC109128 Homo sapi
22	20	1.4	73635	2	AC100991	AC100991 Mus muscu
c 23	20	1.4	79722	2	AC018384	AC018384 Homo sapi
c 24	20	1.4	110000	2	LMFLCHR25_02	Continuation (3 of
c 25	20	1.4	126059	8	AP003197	AP003197 Oryza sat
26	20	1.4	129007	2	AC106094	AC106094 Rattus no
27	20	1.4	138025	9	HSDJ20N4	AL096709 Human DNA
c 28	20	1.4	162501	2	AC097791	AC097791 Rattus no
29	20	1.4	166048	9	AC013725	AC013725 Homo sapi
c 30	20	1.4	167449	2	AC025073	AC025073 Homo sapi
31	20	1.4	172538	2	AC041016	AC041016 Homo sapi
c 32	20	1.4	173333	2	AC025282	AC025282 Homo sapi
c 33	20	1.4	174894	3	AC007984	AC007984 Drosophil
c 34	20	1.4	175347	2	AC009099	AC009099 Homo sapi
c 35	20	1.4	177241	2	AC102960	AC102960 Rattus no
c 36	20	1.4	187388	2	AC094898	AC094898 Rattus no
c 37	20	1.4	188216	2	AC092711	AC092711 Mus muscu
c 38	20	1.4	225092	3	AE003758	AE003758 Drosophil
c 39	20	1.4	251700	1	AP000062	AP000062 Aeropyrum
40	19	1.3	578	6	AX067043	AX067043 Sequence
c 41	19	1.3	600	6	AX053400	AX053400 Sequence
c 42	19	1.3	1015	6	AX053304	AX053304 Sequence
43	19	1.3	1251	10	AF233226	AF233226 Mus muscu
44	19	1.3	1887	6	AX122669	AX122669 Sequence
45	19	1.3	2010	6	AX067041	AX067041 Sequence

ALIGNMENTS

RESULT	1					
AX076845						
LOCUS	AX076845	Sequence	4 from Patent	1464 bp	DNA	linear
DEFINITION	AX076845	Sequence	4 from Patent	WO0105980.		
ACCESSION	AX076845					
VERSION	AX076845.1	GI:13121519				
KEYWORDS						
SOURCE		Exophiala spinifera.				
ORGANISM		Exophiala spinifera				
		Eukaryota; Fungi; Ascomycota; Pezizomycotina; Chaetothryiomycetes;				
		Chaetothryiales; Herpotrichiellaceae; Exophiala.				
REFERENCE		1 (bases 1 to 1464)				
AUTHORS		Duvick,J.P., Maddox,J., Gilliam,J., Folkerts,O. and Crasta,O.R.				
TITLE		Compositions and methods for fumonisin detoxification				
JOURNAL		Patent: WO 0105980-A 4 25-JAN-2001;				
		Pioneer Hi-Bred International Inc. (US) ; Curagen Corporation (US)				
FEATURES		Location/Qualifiers				
source		1. .1464				
		/organism="Exophiala spinifera"				
		/db_xref="taxon:91928"				
		/note="aldehyde dehydrogenase , coding sequence"				
BASE COUNT		346 a	402 c	406 g	310 t	

ORIGIN									
Query Match		100.0%;		Score 1464;		DB 6;		Length 1464;	
Best Local Similarity		100.0%;		Pred. No. 0;					
Matches 1464;		Conservative 0;		Mismatches 0;		Indels 0;		Gaps 0;	
QY	1	atgggtcttcttcgctgacgaaatacaagagtgaaactcttcatcaacaatgaattcgtctcc	60						
Db	1	ATGGTTCCTTTCGCCTGACGAATACAAGAGTGAACCTCTTCATCAACAATGAATTCGTCCTCC	60						
QY	61	tccaaggggtcccgagagattaaacgctcaagaccccgtaggacgaatccaccgttgccact	120						
Db	61	TCCAAGGGGTCCGAGAGATTAAACGCTCACGAACCCGTGGGACGAATCCACCGTTGCCACT	120						
QY	121	gatgttcaacgtggccaaacgcgcgatgtcgacagtgcaagtagccgcttcggtgcaagcg	180						
Db	121	GATGTTCACTGCTGGCCAAACCGCGCGGATGTCGACAGTGCAGTAGCCGCTTCGGTGCAGGCG	180						
QY	181	gtcaaaaaagggcccatgggaagaagttcacaggtgcacaacgcgcggtgcgtgcatgcttaag	240						
Db	181	GTCRAAAGGGCCCATGGGAAGAAGTTCACAGGTGCACAACGCGCGGCTGCATGCTTAAG	240						
QY	241	ttcgcggacctgcgcgagagaagaacgcgcgagaagctcgtctgagtcgctgcccacc	300						
Db	241	TTCGCGGACCTGCGCGAGAGAAGCGCCGAGAAGCTCGCTCGTCTGGAGTCGCTGCCACC	300						
QY	301	ggtagaccgggtgtcgatgatacactcaatttcgacattccaaacatggtctccgtgtttcgc	360						
Db	301	GGTAGACCGGTGTCGATGATCACTCATTTTCGACATTCCAAACATGGTCTCCGCTGTTTCGC	360						
QY	361	tactatgcaggctgggcccacaagaatcgccggaaagaccttcccgaggacaacggaag	420						
Db	361	TACTATGCAGGCTGGGCCGACAAGATCGCCGGAAGACCTTTCCCGAGGACAACGGCAAG	420						
QY	421	ccgaattggcgttacgagccgatgggggtgtgtcgtggtattgccagctggaacgcgact	480						
Db	421	CCGAATTGGCGTTACGAGCCGATGGGGGTGTGTGCTGTGATTTGCCAGCTGGAACGCGACT	480						
QY	481	ttctcttacgtcggctggaagatagccccccgcttcgcccgcggtcgtcctctcatcttc	540						
Db	481	TTCTCTTACGTGGCTGGAAGATAGCCCCCGCTCGCCGCGGCTGCTCTCATCTTC	540						
QY	541	aaagcctcggagaaatccccgcctggcggttctgggctcgtcctctcttcgcagaagcc	600						
Db	541	AAAGCCTCGGAGAAATCCCGCTGGGGGTTCTGGGCTCGCTCCTCTCTTCGAGAAAGCC	600						
QY	601	ggattccctcctggagtcgtgcagttcctcactgagcacgagtgacgggtgaagcattg	660						
Db	601	GGATTCCCTCCTGGAGTCGTGCAGTTCCTCACTGAGCAGCAGTGACGGGTGAAGCATTG	660						
QY	661	gcgtcgcacatggacattgcgaagatcagcttcaagaatctgtcggcggtgcccgcgccc	720						
Db	661	GCGTCGCACATGGACATTCGGAAGATCAGCTTCACAAGATCTGTTCGGCGGTGGCCGCGCC	720						
QY	721	gtcaagcaagcaaacactcaagtccaacatgaagcgcgctcactctagaaactgggggaaaag	780						
Db	721	GTCAAGCAAGCAACACTCAAGTCCAACATGAAGCGCGTCACTCTAGAACTGGGGGAAAG	780						
QY	781	ccaaccatcgtcttcaacgaagctcctctcgaacgycagtcggyggaatcgggcaaggat	840						
Db	781	CCAACCATCGTCTTCAACGAAGCTCCTCTCGAAGCGGAGTCGGGGGAATCGGCAAGGAT	840						
QY	841	ttctcaaaaattcgggcaaatcttgggtccccccccctcctgtttgctagtgaatgggaaaat	900						
Db	841	TTCTCAAAATTTCGGGGCAAAATTGGGTCCCCCCCCCTCCTGTTTGTAGTCAATGGGGAAT	900						
QY	901	ttagcggagaaaaattccatggagtcgcgtcatggctcatttggaggctgtcagagatggctt	960						
Db	901	TTAGCGGAGAAATTCATGGAGTCCGTCCGTCACTGCTCATTTGGAGGCTGTCAAGATGGCTT	960						
QY	961	ggccagaaccccatggaaaccccaagaggacgcatggtcccttcgtcgacaagtcccagtag	1020						

Db	961	GCCAGAACCCATTGGAACCCCAAGAGGACGCATGGTCCCTTCGTCGACAAGTCCCAGTAC	1020
Qy	1021	gacagagtccttgggttaacattgacgcttggcaaggataaccgcgcagctcctcactggcgtt	1080
Db	1021	GACAGAGTCTTGGGTAACATGACGTTGGCAAGGATACCGCGCAGCTCCTCACTGGCGTT	1080
Qy	1081	ggtagaaaggcgacaaaggattcgcgattgaaccgacgatatttgtcaatcccaaacca	1140
Db	1081	GGTAGAAAGGGCGACAAAGGATTGCGGATTGAACCGACGATATTTGTCAATCCCAACCA	1140
Qy	1141	ggcagcaaaaatttggtttgaggagatctttggccccgtcttgtccattaaagacggttcaag	1200
Db	1141	GGCAGCAAAATTTGGTTTGAGGAGATCTTTGGCCCCGCTCTTGTCATTAAAGACGTTCAAG	1200
Qy	1201	acggaagaaggagccattgagattgccaatgacacgacttatgggctagcctcgggtcatt	1260
Db	1201	ACGGAAGAAGAGGCCATTGAGATTGCCAATGACACGACTTATGGGCTAGCCTCGGTCAAT	1260
Qy	1261	tataccaaatctctcaacaggggtctccggtctcgttcgtcgcgctcgagaccggtggcgctc	1320
Db	1261	TATACCAATCTCTCAACAGGGGTCTCCGCTGTCTCGTCGGCGCTCGAGACCGGTGGCGTC	1320
Qy	1321	tcgatcaacttccccctttatccccgagacacaaaactccgcttggcgcatgaaacaatcg	1380
Db	1321	TCGATCAACTTCCCCCTTTATCCCCGAGACACAAACTCCGTTTGGCGGCATGAACAATCG	1380
Qy	1381	ggctcaggcagagagctaggcggaagaaggctcaaggcgctacttggagcccaagaccatt	1440
Db	1381	GGCTCAGGCAGAGAGCTAGGCGAAGAAGGGCTCAAGCGGTACTTGGAGCCCCAAGACCATT	1440
Qy	1441	aatatccacgtcaacatagagtga	1464
Db	1441	AATATCCACGTCACATAGAGTGA	1464
RESULT 2			
ABY17825			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
University, Section MGIM, Dreijenlaan 2, NL- 6703 HA Wageningen, NETHERLANDS			
FEATURES			
source			
1. .4158			
/organism="Agaricus bisporus"			
/strain="Horst H39"			
/db_xref="taxon:5341"			
join(<532. .650,716. .729,787. .1170,1234. .1439,1494. .1657,1713. .1881,1934. .2123,2190. .2302,2361. .2455,2516. .>2564)			
mrna			
/gene="alda"			
532. .2564			
/gene="alda"			
<532. .650			
/gene="alda"			
/number=1			
join(532. .650,716. .729,787. .1170,1234. .1439,1494. .1657,1713. .1881,1934. .2123,2190. .2302,2361. .2455,2516. .>2564)			
CDS			

/gene="alda"
/EC_number="1.2.1.3"
/codon_start=1
/product="putative aldehyde dehydrogenase (NAD+)"
/protein_id="CAA76875.1"
/db_xref="GI:3355324"
/db_xref="SPTREMBL:O74187"
/translation="MPSIFTHQWDTPVYKGSTSINTGLFINGEFVDGVKNTTIDVVNP
ANGKLITKISEATEADIDIAVEAAHKAFETTWGLNCSGSKRGDMLYKLAQLMKKNIDD
LSALEALDNGKTELWAKSVDLSLSSISIKHYAGWADKNFGQVIETDEKKLTYSRHEPI
GVVGQIIPWNFLLMAWKIGPALATGNCIVLKPSEFTPLSALRMCALIQEAGFPFPGV
VNVVTGYGSTTGOATSSHMKIDKVAFTGSTLVGRKVMEEAAKSNLKNVTLELGGKSPV
VIFDDADLEQSVNWTAGHLEFNHGOACCACTRIFFVQEGYIDKFLQKFTDKIKEIKLGD
PFGLGIDQGPQVQIQYDRIMSYIESGRAEGATVHVGGERHNEGFIQPTIFTDTTP
DMKIVKEEIFGPVGAVIKFKDGKEVIKQANDSNYGLAAAVFSQDINKAIETAHAFKAG
TAWVNCANTIDAGVPFGGYKQSGIGRELGEYALHNYTNVKAHVNLNWKM"
551. .715
/gene="alda"
/number=1
716. .729
/gene="alda"
/number=2
730. .786
/gene="alda"
/number=2
787. .1170
/gene="alda"
/number=3
1171. .1233
/gene="alda"
/number=3
1234. .1439
/gene="alda"
/number=4
1440. .1493
/gene="alda"
/number=4
1494. .1657
/gene="alda"
/number=5
1658. .1712
/gene="alda"
/number=5
1713. .1881
/gene="alda"
/number=6
1882. .1933
/gene="alda"
/number=6
1934. .2123
/gene="alda"
/number=7
2124. .2189
/gene="alda"
/number=7
2190. .2302
/gene="alda"
/number=8
2303. .2360
/gene="alda"
/number=8
2361. .2455
/gene="alda"
/number=9
2456. .2515
/gene="alda"
/number=9
2516. .>2564
/gene="alda"
/number=10
<3094. .3140
/gene="echa"
/number=1

intron

exon

intron

exon

intron

exon

intron

exon

intron

exon

intron

exon

intron

exon

intron

exon

intron

exon

intron

exon

intron

exon

mRNA
gene
CDS

intron

exon

intron

exon

intron

exon

misc_feature

BASE COUNT 1049 a 971 c 955 g 1183 t
ORIGIN

Query Match 1.6%; Score 24; DB 8; Length 4158;
Best Local Similarity 100.0%; Pred. No. 0.15;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 362 actatgcaggctggcgacaaga 385
|||||
Db 1048 ACTATGCAGGCTGGCGACAAGA 1071

RESULT 3
AP003048
LOCUS AP003048 157201 bp DNA linear PLN 22-MAY-2001
DEFINITION Oryza sativa genomic DNA, chromosome 1, PAC clone:P0710A02.
ACCESSION AP003048
VERSION AP003048.2 GI:14164403
KEYWORDS
SOURCE
ORGANISM
Oryza sativa (cultivar:Nipponbare) DNA, clone:P0710A02.
Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
1 (sites)
Sasaki,T., Matsumoto,T. and Yamamoto,K.
Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
clone:P0710A02
Published Only in DataBase (2000) In press
2 (bases 1 to 157201)
Sasaki,T., Matsumoto,T. and Yamamoto,K.
Direct Submission
Submitted (13-DEC-2000) Takuji Sasaki, National Institute of
Agrobiological Resources, Rice Genome Research Program; Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan

(E-mail:tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
Tel:81-298-38-7441, Fax:81-298-38-7468)
On May 21, 2001 this sequence version replaced gi:11862978.
Genes were predicted from the integrated results of the following:
GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as SplicePredictor
(October 1998 version). The genomic sequence was searched against
NCBI NonRedundant Protein database, nr
(ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at
RGP. Protein homologies of the coding regions were searched against
NCBI NonRedundant Protein database with BLASTP2.0. ESTs represent
the identified cDNA sequences using BLASTN 2.0 with the
corresponding DDBJ accession no. and RGP clone ID.
A gene with identity or significant homology to a protein is
classified based on the protein name to indicate the homology level
such as same name, 'putative-' and '-like protein'. A gene without
significant homology to any protein but with EST homology (covering
almost the entire length of partial sequence) is classified as an
'unknown' protein. A gene predicted with a gene prediction program
is classified as a 'hypothetical' protein.
The orientation of the sequence is from SP6 to T7 of the PAC clone.
Detailed information on overlap and assembly quality together with
annotation of this entry is available at
http://rgp.dna.affrc.go.jp/Genomeseq.html.

FEATURES
Source

1..157201
/organism="Oryza sativa"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/chromosome="1"
/clone="P0710A02"
join(4865..5500,5607..5668,5918..6014,6862..7143,
7382..7435,7532..7678)
/gene="P0710A02.1"
join(4865..5500,5607..5668,5918..6014,6862..7143,
7382..7435,7532..7678)
/gene="P0710A02.1"
/note="contains ESTs AU065254(E60905),AU065254(E60905)
unknown protein"
/codon_start=1
/protein_id="BAB55803.1"
/db_xref="GI:14164404"
/translation="MPSDDEVREAGEVMILHFPEAAKSDNLFDKIHKNFRSPAHVFE
GKRQIEKARSWSKTCGDVKSIFQGDGAGSEIAEKGRRDDHDATEKSKKDYTEAEAKEF
LEEVRLIHRTFDKAEENFLKKDMPVHLSIDGPIDVAWTENELRRSREHGLDDPSWGFL
TERSPKFPPTNNNQNTLIENAYMTFSECDEYSGSPDHQEEEMNEESVGKYVTSARA
GVFFCOTASEFEMRIIPVFESSKKDVLSSFDLPDLPELPHRLIKHLIAYHESGHAI
VALHTQGAHPHQATVPLPHGLSLGMVTQVASPGNTSISOKILARIDVCFGGAAEEL
LFGEHTVTAGAKNDLFTAKELAQLVLKLLGESRERVKRLKKHEMQLQAVANALLEK
EILNADEIHKVQPYKREPOITFQDKELAVNWS"
complement(join(10619..10756,10904..11062,12001..12496,
12591..12733,13399..13580,13819..14023,14428..14565))
/gene="P0710A02.2"
complement(join(10619..10756,10904..11062,12001..12496,
12591..12733,13399..13580,13819..14023,14428..14565))
/gene="P0710A02.2"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAB55804.1"
/db_xref="GI:14164405"
/translation="MIKWLDLIPRQTMKQSVLANHQPCFGVNSYDHHKXRDERYLL
GMEISRKWLATYJRNAEHTAEAYKEFVASGYDYAQLLRVPLRLFPKTSYHCIKKSIYC
DHHEYLTFTTSETASTIGMNEPNOTIQIFSLKSSNESYPISVYGIFAVRDELDRLRN
YVFNRTDNPVLIEQFINRDLRLSNKKYLEVPVPHGTSFIVESNNAHPITOLDPTIG
NDLVPPCRGMYLILSFALFEVDLWVKAESGGSDDKQLLSEYAEIENRGLKDPMIYGR
IPSDRCLLDVDCMFLNLSVEAVIQVFTKDDSDNPHRVRTAFSSGFDHEIVLYDDRLC
KXGKLFQNVAVAKSEKLVIRLEFEGSTFRWTFQDGAVAAVSSPDDSVSKLEFDLWKA
RNDLKFOGKIKEPTQVCEAKMIKTYTNFLSNLMOQEDLDEEDTNRMRMGGAASISTG
NGDCQILEVHFLTDSAILARTLARETFKKNQGIGA"
join(18601..18688,19025..19173,19275..19365,19901..20054,
20134..20182,20516..20630,20751..20842,20938..21085,
21515..21579,21654..21751,22315..22420,22959..23096)
/gene="P0710A02.3"
join(18601..18688,19025..19173,19275..19365,19901..20054,

20134..20182,20516..20630,20751..20842,20938..21085,
21515..21579,21654..21751,22315..22420,22959..23096)
/gene="P0710A02.3"
/codon_start=1
/product="putative protein kinase AFC3"
/protein_id="BAB55805.1"
/db_xref="GI:14164406"
/translation="MPEPATIQGSNLSAILSWSSSPDALKSGFLSLSLSPLSPL
SPIESESMESSRSRKRAQDWDGAGAPPPPAERAVVARGGASPPWRDDDDRDGHYVFDL
GENLNRRCTFGRVLECWDRETHHEYVAIKVRSIRKYRDAAMIEDVLRNLAENERYRS
LCVQIQGWFDYRNHICIVFEKLGPSLYDFLKRNRYPFPVELVREFGRQLLESVAYMH
ELRLHTDLKPENILLVSSEYIRVPGSKNSQDEMFKCLPKSSAIKILIDFGSTAFDN
QEHSSIVSTRHYRAPEIILGLGWSFPCDLWSVGCIIVELCSGALFQTHENLEHLAMM
ERVGLPLPEHMIKASSSAQKYFRRGARLNWPEGAVSRESIRAVKKLDRLKDLVARKA
DHSRAVLADLLYGLKPEPSERLTAQEALDHPFFRNTT"
join(25678..25783,32101..32243,32405..32878,33109..33282,
33389..33526,33648..33785,33910..33971,34069..34342)
/gene="P0710A02.4"
join(25678..25783,32101..32243,32405..32878,33109..33282,
33389..33526,33648..33785,33910..33971,34069..34342)
/gene="P0710A02.4"
/note="contains ESTs D24173(R1467),AU057957(S20436)"
/codon_start=1
/product="cytosolic aldehyde dehydrogenase"
/protein_id="BAB55806.1"
/db_xref="GI:14164407"
/translation="MAAANGDSKGFEPKLEIKFTKLFINGREFVDVAVSGKTFETRDP
RTGEVIAKIAEGDKADIDLAVKAAREAFDHGPPWMSGFARGRIILHKFADLVEQHVVE
LAALDTVAGKLFANGKLVLDIPGGANLLRYAGAADKVHGETLKMARPCGYTLKEPV
GVVGHIVPNYPTTMTFFKASPALAAAGCTMVVKPAEOTPLSALFYAHLAKLAGVPDGV
LNVVPGFGPTAGAAISSHMDIDKVSFTSGSTEVRGLVMEAAKSNLKPVSLELGGKSPV
IVFDDADLDTAVNLVHMASYTNKEICVAGSRIYVQEGIYDAFVKKATEMAKKSVMYGD
PFNPRVHQGPQIDKEQYEKILKYIDIGKREGATLVTTGGKPCGSENGYVIEPTIFTDVKE
EMSI AQEEIFGPMALMKFTVEEAIQKANSTRYGLAAGIVTKNIDVANTVSRISIRAG
AIWICYLGFDPDVPFGYKMSGFGKDMGMDALEKYLHTKAVVTPLYNTPLW"
complement(39757..39978)
/gene="P0710A02.5"
complement(39757..39978)
/gene="P0710A02.5"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAB55807.1"
/db_xref="GI:14164408"
/translation="MHRKEIINHNEIDGDNEDVVEAKLVKPCMAERDWMKRLDMAV
GVLRYRRGSATNDMPDARGDRRSATARRL"
join(42079..42199,45522..45664,45785..46168,46289..46378,
46469..46642,46767..46904,47043..47180,47253..47314,
47427..47700)
/gene="P0710A02.6"
join(42079..42199,45522..45664,45785..46168,46289..46378,
46469..46642,46767..46904,47043..47180,47253..47314,
47427..47700)
/gene="P0710A02.6"
/note="contains EST AU164502(R0133)"
/codon_start=1
/product="putative cytosolic aldehyde dehydrogenase"
/protein_id="BAB55808.1"
/db_xref="GI:14164409"
/translation="WGSTGDCGNGKANAGGGGLVVPKIKFTKLFINGEFVDAASGKTF
KTRDPRTGDVLAHTAEADKADVDLAVKAAREAFEHGKWPMSGYERSRVNMKLDLVE
QHADELAALDGADAGKLLTLGKIIDMPAAQMMRYAYAGAAKIHGESLRVAGKYQGYT
LREPIGVGVVPIPNWPTMMFFLKVSPALAAAGCTIVVKPAEQTPLSALYYAHLAKLAG
VPDGVINVVPGFGPTAGAAALSSHMDVDSVAFVTSAGIEGRAIMESAARSNLKNVSLELG
GKSPMIVFDDADVDMAVLSLSSLAFFNKGEICVAGSRVYVQEGIYDEFFVKVAEAKN
WKVGDPFDDAATMGPOVDKVQFERYLVKYEIGKNEGATLLTGKPTGDKGYVIEPTIF
VDVKEEMTIAQEEIFGPMVMSLMKFTVEEAIKANCTKYGLAAGIVTKNLNIANMYSR
SVRAGTVWVNCYAFDPDPAFGGYKMSGFGRQGMVAMDKYLQVKTVITAVPDSPTW"
complement(52056..52268)
/gene="P0710A02.7"
complement(52056..52268)
/gene="P0710A02.7"
/note="hypothetical protein"
/codon_start=1

misc_feature

protein_id="BAB55809.1"
db_xref="GI:14164410"
translation="MASLKLRCVLLSTVFFSGLMQLSMAQAGSPAATATATTARV
IDVKAVDQAVAYLLMLAALFVTVLAH"
join(53801..55766,65709..69632)
/gene="P0710A02.8"
/note="probably inactive due to frameshift in CDS
pseudogene, similar to Oryza sativa Mutator-like
transposase"
/pseudo
join(53801..55766,65709..69632)
/gene="P0710A02.8"
complement(join(62257..62378,63378..63507))
/gene="P0710A02.9"
complement(join(62257..62378,63378..63507))
/gene="P0710A02.9"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAB55810.1"
/db_xref="GI:14164411"
/translation="MDISRTNEGYTSCGPVVMWSHRAGVLLLGHSCLPVEVPNAV
LDGQLCEEEDVNCPLPKLWKTIGSASEVEEAAPDKQE"
join(78407..78539,79119..79243,80627..80731)
/gene="P0710A02.10"
join(78407..78539,79119..79243,80627..80731)
/gene="P0710A02.10"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAB55811.1"
/db_xref="GI:14164412"
/translation="MWAREANQDSAPPLAGPVQPIFGMTVVLLIPMTVVVAISTGGNRH
NRPLHYKKKLTIISSNAADLNSRPEVEQETRGPERGREDWRMKMKYSSTRIHDPASS
KNIHDPQSGFSKLTSVNR"

gene

CDS

Query Match 1.5%; Score 22; DB 8; Length 157201;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 507 cccgcgcctgcgcgcgctgc 528
|||||
Db 46078 CCCGCCCTCGCCGCGGCTGC 46099

RESULT 4
RME603643
LOCUS
DEFINITION Rhizobium meliloti (Sinorhizobium meliloti) 1021 complete pSymb;
segment 2/6.
ACCESSION AL603643 AL591985
VERSION AL603643.1 GI:15140164
KEYWORDS
SOURCE Sinorhizobium meliloti.
ORGANISM Sinorhizobium meliloti
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Rhizobiaceae; Sinorhizobium.
REFERENCE 1 (bases 1 to 260050)
AUTHORS Finan,T.M., Weidner,S., Wong,K., Buhrmester,J., Chain,P.,
Vorholter,F.J., Hernandez-Lucas,I., Becker,A., Cowie,A., Gouzy,J.,
Golding,B. and Puhler,A.
TITLE From the Cover: The complete sequence of the 1,683-kb pSymb
megaplasmid from the N2-fixing endosymbiont Sinorhizobium meliloti
JOURNAL Proceedings of the National Academy of Sciences of the United
States of America. 98 (17), 9889-9894 (2001)
PUBMED 11481431
REMARK epub ahead of print
REFERENCE 2 (bases 1 to 260050)
AUTHORS Weidner,S.
TITLE Direct Submission
JOURNAL Submitted (07-JUN-2001) Weidner S., Universitaet Bielefeld,
Biologie IV (Genetik) Universitaetstr 25, D-33615 Bielefeld,
Germany
COMMENT Submitted on behalf of Universitaet Bielefeld, Biologie IV
(Genetik) Universitaetstr 25, D-33615 Bielefeld, Germany and

Department of Biology, McMaster University, 1280 Main Street West,
Hamilton, Ontario, L8S 4K1 Canada
mailto:Stefan.Weidner@Genetik.Uni-Bielefeld.DE
pEXO, pSymb.

FEATURES

source

Location/Qualifiers
join(1..745,join(1805..124686,125855..260050))
/organism="Sinorhizobium meliloti"
/plasmid="pSymb"
/strain="1021"
/db_xref="taxon:382"
/focus
complement(120..281)
/gene="Smb20302"
complement(120..281)
/gene="Smb20302"
/function="MISCELLANEOUS; Hypothetical/Global homology"
/note="Product confidence : putative
Gene name confidence : hypothetical"
/codon_start=1
/transl_table=11
/evidence="not_experimental
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAC48692.1"
/db_xref="GI:15140165"
/translation="MSVLISILITFLVVLVLYLVNRLPLDGRTKQIVQVIVIIIGIL
SLRLYLVF"
478..705
/gene="Smb20303"
478..705
/gene="Smb20303"
/function="MISCELLANEOUS; Unknown"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical"
/codon_start=1
/transl_table=11
/evidence="not_experimental
/product="HYPOTHETICAL PROTEIN"
/protein_id="CAC48693.1"
/db_xref="GI:15140166"
/translation="MSVRFSDLAIDDBEILMLRRVHVACSSRELEPESEHGAEGLGKM
LFHLYRQGVRSSELALQMMLTAGAAPAPEAG"
complement(746..1804)
/note="this element seems to be partial or inactive"
/organism="Sinorhizobium meliloti"
/insertion_seq="ISRM2011-2/ISRM11 OR Smb21616"
/db_xref="taxon:382"
complement(762..1367)
/gene="TRM2011-2C OR Smb20304"
complement(762..1367)
/gene="TRM2011-2C OR Smb20304"
/function="ELEMENTS OF EXTERNAL ORIGIN; Transposon-related
functions"
/note="Product confidence : probable
Gene name confidence : putative"
/codon_start=1
/transl_table=11
/evidence="not_experimental
/product="probable ISRM2011-2 transposase protein,
C-terminal portion"
/protein_id="CAC48694.1"
/db_xref="GI:15140167"
/translation="KKTIVASERERPDVARHRARWLKHCPCIDPARLVFIDEIWTWKTN
MAPLRGWAPRGERLVGYAPFGHWNTMTFVAALRADRVSAFILDGPINGERFRIVQQ
VLVPELKAGDIVILDNLGSHKQGEIRAIRKAGARLFFLPKYSFDLNPICKLFAKIKH
WLREAQARSDAIDELRHILQAVTPQCAAYFKEAGYERA"
1002..1373
/gene="TRM2011-2 OR Smb20306"
1002..1373
/gene="TRM2011-2 OR Smb20306"
/function="ELEMENTS OF EXTERNAL ORIGIN; Transposon-related
functions"
/note="Product confidence : putative
Gene name confidence : putative"

gene

CDS


```
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative TRM2011-2 protein"
/protein_id="CAC48695.1"
/db_xref="GI:15140168"
/translation="MGAEIVQDDVAGFEFRHQLLDIDAFAIDRAIEDKGSADAV
GPEGDKGHGVPMAEGVAHQSFARPPAPQRRHRLRPGLDENEAGGINSGLVLP
SCPVAGDVGPPLPFAGDORLFL"
complement(1301..1708)
/gene="TRM2011-2N OR Smb20305"
complement(1301..1708)
/gene="TRM2011-2N OR Smb20305"
/function="ELEMENTS OF EXTERNAL ORIGIN; Transposon-related
functions"
/note="Product confidence : probable
Gene name confidence : putative"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="probable ISRM2011-2 transposase protein,
N-terminal portion"
/protein_id="CAC48696.1"
/db_xref="GI:15140169"
/translation="MARFSDNLRVRVDAVTGEGLSCLRAAKRFGIGISTADNVRR
FRETCSAAPQGMGHPKRLKSGPHRAWLLCRERDFTLHGLVAELSERGLKVDYRAV
WTFVHEGLSYKKRWSPANGSGPTSPATGHDG"
complement(1801..2448)
/gene="Smb20307"
complement(1801..2448)
/gene="Smb20307"
/function="MISCELLANEOUS; Hypothetical/Partial homology"
/note="Product confidence : putative
Gene name confidence : hypothetical"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative dihydroxyacetone kinase protein"
/protein_id="CAC48697.1"
/db_xref="GI:15140170"
/translation="MNSSAVRRLIEVFAISSAIEAERDHLSELDGAIGDADHGITM
AIGPSAVTGELARTDAGSAGLSGVMTAAAFNLAVGASTGPLYATAFRRAAQALEGR
DTLDIEACSLLIQAVSGIGERGKGQGDKTMLDVWLPAAADAAQAVAAKSTAAFWA
DVTAAENGASATRSNVATKGRAARVGRSLGHMDPGAASAVLLIRAMAQTLREG"
complement(2527..4614)
/gene="Smb20312"
complement(2527..4614)
/gene="Smb20312"
/function="MISCELLANEOUS; Not classified regulator"
/note="Product confidence : putative
Gene name confidence : hypothetical"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative transcriptional regulator protein"
/protein_id="CAC48698.1"
/db_xref="GI:15140171"
/translation="MTIKTPSARKGAANRSGTIDASDAIPRYGDDPYVWACWLYYED
GMTQGEIAEAMGVSRATVNSYLADAREKIVNISIEPARLASLTIAQELKRHFGLTDC
LVVPSDDNARPLIDRLGAAGAQAQALPKLSGDTLAVAWGRVYLSVGHAGIASLQDMT
VVOATGGTTASFAVTPELCASAVARAI SARCVNITAPAVVKSPELLHMLLEPLVQEQ
FATLARANRVLFGLSSLRPNSTIHTSGFEFSVSLQDYLAAGAVGVWAGRFIDRGRPV
AGPLDHRVTGIVSLDLRLKIGTRIAVAGGFDKVPALLAALRGVYVNVLTIDATGHGIL
RADGVTDALDAKLNHRPKLVPTSSRYTHVKKFLNPNVEVVEMLDGAVKAHEIYLQPI
NGSHRALVARNGPFGKVLVIGGGTGHEPCFLGYVGKGLADAVAVGNIFSSPPDPPI
VQCAEASGGEGVLFVYGNAGDVMEFMAEIAEEKGIKVTVLTDDVASSPVEDR
EGRRGVAGNFFIFKIAAACDRLGLEACEAVTRKANLRTYTVGVALEPGSMPTQRRH
NFEIGPDDMEVGMGIHGPVTRERIRSADEITDSIMDRIFKEMKTAPGERVAVLVNS
FGATPLMELYILFRRVQORLAADKILIEANWIGHYCTSLDMAGASITILHLDQELSDL
LHPCETAFLKN"
complement(4728..5369)
/gene="Smb20313"
complement(4728..5369)
```

```
/gene="Smb20313"
/function="Small Molecule Metabolism; Central intermediary
metabolism; misc. glycerol metabolism"
/note="product confidence : putative
Gene name confidence : hypothetical"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative dihydroxyacetone kinase protein"
/protein_id="CAC48699.1"
/db_xref="GI:15140172"
/translation="MQTFENARAGDIVLTADRIENRAYLSEIDKIGGDHGVNMA
KGFGLAAERLKGKDALAEALDTLGTVMTEIGGSMGPLYGVMTFPAEKIEGADAID
AAYSRMLHAGLEGIRSIGSAKVGDKTLDTLVPADVAFDAATADGPPFAEALDALVS
AAEAGRDSTRDLIARIGRASRLGERSLGLVDAGATSCAIIILRELSHGARALA"
complement(5380..6372)
/gene="Smb20314"
complement(5380..6372)
/gene="Smb20314"
/function="Small Molecule Metabolism; Central intermediary
metabolism; misc. glycerol metabolism"
/note="Product confidence : putative
Gene name confidence : hypothetical"
/codon_start=1
/transl_table=11
/evidence=not_experimental

Query Match 1.5%; Score 22; DB 1; Length 260050;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 gccgcttcggtgcaggcggtca 184
|||||
Db 175408 GCCGCTTCGGTGCAGCGGTCA 175429

RESULT 5
LOCUS HSU62769 3897 bp mRNA linear PRI 15-AUG-1997
DEFINITION Human oxytocinase variant 2 mRNA, complete cds.
ACCESSION U62769
VERSION U62769.1 GI:2209277
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 3897)
AUTHORS Laustsen,P.G., Rasmussen,T.E., Petersen,K., Pedraza-Diaz,S.,
Moestrup,S.K., Gliemann,J., Sottrup-Jensen,L. and Kristensen,T.
TITLE The complete amino acid sequence of human placental oxytocinase
JOURNAL Biochim. Biophys. Acta 1352 (1), 1-7 (1997)
MEDLINE 97320624
REFERENCE 2 (bases 1 to 3897)
AUTHORS Laustsen,P.G., Rasmussen,T.E., Petersen,K., Moestrup,S.,
Gliemann,J., Sottrup-Jensen,L. and Kristensen,T.
TITLE Direct Submission
JOURNAL Submitted (02-JUL-1996) Department of Molecular and Structural
Biology, Aarhus University, Langelandsgade 140, Aarhus 8000 C,
Denmark
FEATURES
source Location/Qualifiers
1 .3897
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="placenta"
359..3394
CDS
/function="inactivates oxytocin and vasopressin"
/note="cysteine aminopeptidase; vasopressinase; leucine
aminopeptidase"
/codon_start=1
/product="oxytocinase splice variant 2"
/protein_id="AAB66673.1"
/db_xref="GI:2209278"
```

/translation="MIENSMFEEEPDVVDLAKPECLHPLEPDEVEYEPGRSRLLVRLG
GEHEMEDEEDYESAKLLGMSFMNRSLRNSATGYRQSPDGACSVPSARTMVVCAF
VIVVAVSIMVTYLLPRCTFTKEGCHKKNQSIGLIOPFATNGKLFPPAQIRLPFAVVP
LRYELSLHPNLTSMTPRGSVTISVQALQVTWNIILHSTGHNISRVTFMSAVSSQEQOA
EILEYAHGQIAIIVAPEALLAGHNYTLKIEYSANISSYYGYGFSYTDSENEKKYFA
ATQFEPLAARSAPFCDEPAFKATFIKIIRDEQYTALESNNPKKSSVLDGGLVQDEF
SSVKMSTYLVAFIVGEMKNLSQDVNGTIVSIYAVPENIGQVHYALETTVKLLERFQN
YFEIQYPLKKLDLVAIPDFEAGAMENWGLTTFREETLIYDSNTSSMADRKLVTXIIAH
ELAHQWFGNLTVMKWNDLWLNNEGFAFMFEYFSLEKIFKELSSYEDFLDARFKTMKKD
SLNSSHPISSSVQSSEQIEEMFDSLXFKGSSLLMLKTYLSEDFVQHAVVLYLHNHS
YASIQSDDLWDSFNEVTNQTLDVKRMMKWTWLOKGFPLTVQKKGKELFIOQERFPLN
MKPEIQPSDTSYLWHPISYVTEGRNYSKYQSVSLLDKKSGVINLTETEVLVWVKYNINM
NGYIIVHYADDDWEALIHOLKINPYVLSDKDRANLINNIFELAGLCKVPLKRAFDLIN
YLGNEHTAPITEALFQTDLIYNLLEKLGMDLASRLVTRVFKLLQNIQOQTWTDG
TPSMRELRSALLEFACTHNLGNCSTTAMKLFDDWMASNGTQSLPTDVMTTVFKVGAKT
DKGWSFLLGKYISIGSEAEKNKILEALASSEDEVKLYWLMKSSSLNGDNFRTQKLSFII
RTVGRHFPGHLLAWDFVKENWNKLVQKFLGYSYTIONIVAGSYTLFSTKTHLSEVQAF
FENQSEATFRLRCVQEALEVIQNLQNMENKLNKSLTWL"

BASE COUNT 1142 a 787 c 846 g 1122 t
ORIGIN

Query Match 1.4%; Score 21; DB 9; Length 3897;
Best Local Similarity 100.0%; Pred. No. 7.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 525 ctgctccttcattcattcaagc 545
|||||
Db 307 CTGCTCCTTCATCTCAAGC 327

RESULT 6
HSA131024
LOCUS HSA131024 4226 bp DNA linear PRI 15-JUN-2000
DEFINITION Homo sapiens otase gene, exon 1-alpha.
ACCESSION AJ131024
VERSION AJ131024.1 GI:6468557
KEYWORDS otase gene; oxytocinase/insulin-responsive aminopeptidase.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4226)
AUTHORS Rasmussen,T.E., Pedraza-Diaz,S., Hardre,R., Laustsen,P.G.,
Carrión,A.G. and Kristensen,T.
TITLE Structure of the human oxytocinase/insulin-regulated aminopeptidase
gene and localization to chromosome 5q21
JOURNAL Eur. J. Biochem. 267 (8), 2297-2306 (2000)
MEDLINE 20223264
REFERENCE 2 (bases 1 to 4226)
AUTHORS Rasmussen,T.E.
TITLE Direct Submission
JOURNAL Submitted (23-NOV-1998) Rasmussen T.E., Department of Molecular and
Structural Biology, University of Aarhus, DK-8000 Aarhus C,
Jylland, DENMARK
FEATURES
source Location/Qualifiers
1. .4226
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/map="5q21"
295. .546
repeat_region /rpt_family="Alu"
/rpt_unit=295. .546
/rpt_type=DISPERSED
1275. .1470
repeat_region /rpt_family="Alu"
/rpt_unit=1275. .1470
/rpt_type=DISPERSED
3412. .3737
exon /gene="otase"
/note="number 1-alpha"
/usedin=AJ131023:otase_mRNA

/usedin=AJ131023:otase_CDS
3412. .3737
gene /gene="otase"
3809. .3990
repeat_region /rpt_family="Alu"
/rpt_unit=3809. .3990
BASE COUNT 1093 a 824 c 831 g 1466 t 12 others
ORIGIN

Query Match 1.4%; Score 21; DB 9; Length 4226;
Best Local Similarity 100.0%; Pred. No. 7.5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 525 ctgctccttcattcattcaagc 545
|||||
Db 3709 CTGCTCCTTCATCTCAAGC 3729

RESULT 7
AB008831
LOCUS AB008831 10498 bp DNA linear BCT 05-JUN-1998
DEFINITION Acinetobacter sp. genes for aniline dioxygenase subunits and
reductase component, partial and complete cds.
ACCESSION AB008831
VERSION AB008831.1 GI:2627146
KEYWORDS 4-oxalocrotonate decarboxylase; 4-hydroxy-2-oxovalerate aldolase;
acetoaldehyde dehydrogenase; 2-oxopent-4-dienoate hydratase;
2-hydroxyomuconic semialdehyde hydrolase; 2-hydroxyomuconic
semialdehyde dehydrogenase; catechol 2,3-dioxygenase; small
ferredoxin-like protein; aniline dioxygenase reductase component;
aniline dioxygenase beta-subunit; aniline dioxygenase
alpha-subunit.
SOURCE Acinetobacter sp. (strain:YAA) DNA.
ORGANISM Acinetobacter sp.
Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
Acinetobacter.
REFERENCE 1 (bases 1 to 10498)
AUTHORS Takeo,M.
TITLE Direct Submission
JOURNAL Submitted (13-NOV-1997) Masahiro Takeo, Himeji Institute of
Technology, Department of Applied Chemistry; 2167 Shosha, Himeji,
Hyogo 671-22, Japan (E-mail:takeo@chem.eng.himeji-tech.ac.jp,
Tel:0792-67-4893, Fax:0792-67-4891)
REFERENCE 2 (sites)
AUTHORS Takeo,M., Fujii,T., Takenaka,K. and Maeda,Y.
TITLE Cloning and sequencing of a gene cluster for the meta-cleavage
pathway of aniline degradation in Acinetobacter sp. strain YAA
JOURNAL J. Ferment. Bioeng. 85, 514-517 (1998)
REFERENCE 3 (sites)
AUTHORS Fujii,T., Takeo,M. and Maeda,Y.
TITLE Plasmid-encoded genes specifying aniline oxidation from
Acinetobacter sp. strain YAA
JOURNAL Microbiology (Reading, England) 143 (Pt 1), 93-99 (1997)
MEDLINE 9717776
PUBMED 9025282
FEATURES
source Location/Qualifiers
1. .10498
/organism="Acinetobacter sp."
/strain="YAA"
/db_xref="taxon:472"
1. .276
/gene="atdA3"
<1. .276
/gene="atdA3"
/codon_start=1
/transl_table=11
/product="aniline dioxygenase alpha-subunit"
/protein_id="BAA23550.1"
/db_xref="GI:2627147"
/translation="ELNAIRMRTQEDFPIMGEVDDVANPESCQEGLETMPETIEWIDFS
RHMNEGENACYQDVIOHKPTSEIHSRHYFDTWLQLSAVNKENQSEV"

gene	278. .874 /gene="atda4" 278. .874 /gene="atda4" /codon_start=1 /transl_table=11 /product="aniline dioxygenase beta-subunit" /protein_id="BAA23551.1" /db_xref="GI:2627148" /translation="MNNKDLISIKIVDPSAVNLAYYQEI KQYSDYFWNLNLGEPALD HKINMFLTKEARLLDQQCFDEWLFLEDDGCIWIPGSPAAPASEATYEFHDIRRLK DRIVRLQTGFAYSQIPVSKTNRI LGAPEVWAVPGSSEGELVRTSFIVFESRDGKSQVL SGWYGYVIIKDNDELKIKMKQINLNDCLSPQGNSSF" 885. .1895 /gene="atda5" 885. .1895 /gene="atda5" /codon_start=1 /transl_table=11 /product="aniline dioxygenase reductase component" /protein_id="BAA23552.1" /db_xref="GI:2627149"	CDS	4187. .5647 /gene="atdC" /codon_start=1 /transl_table=11 /product="2-hydroxymuconic semialdehyde dehydrogenase" /protein_id="BAA23556.1" /db_xref="GI:2627153" /translation="MKDIKHYINGQYVGSASGKLFNDNVNPANGKLLSKVHEAGREEVD AAVNAAKAALRGPGWKMTLEERTSILHKVADGINARFDEF LKAECLDTGPKPKSMASHI DIPRGAAHFV FADMIKINVPTESFEMPTPDGTGALNYAVRRPKGVIGVISPWNLP LLL MTWKVGPALACANTVVVKPSEETPTTTALLGEVYKEAGIPTGVYVNVHGF'GGDSAGAF LTEHPQVDGFT'FTGETVPEVIMKAAKGIRDISLELGKNAALYFADCDMDKAIEGT MRSAFANCGQVCLGTERVYVERSIFDEFVGR LKAAEGLKIGPPDDAEANLGLVSLK HREKVL SY YQKAVDDGATVVTGGGVPEMPEELAGGANVQPTIWTGLPKDSAVITDEIF GPCVHVCPFDTEEEAIELANSLPYGLASAIWSENITRAHRVAGQIEAGIIWVNSWFLR DLRTPFGGSKSGIGREGVHSL EFTYTEMKNICVKL" 5655. .6512 /gene="atdD" 5655. .6512 /gene="atdD" /codon_start=1 /transl_table=11 /product="2-hydroxymuconic semialdehyde hydrolase" /protein_id="BAA23557.1" /db_xref="GI:2627154" /translation="MTAPQNSPEIGREITACGYRTNVHDHGADNAKSDVAVMMIHGSG PGVTAWANWRLVIPELAKNRRVLADPMLGFGYTERPEDNTYNNRERWVKHAVGVMDLGL LEQVDLVGNSFGGGLALALAEHPQRRRLVLMGVSQVSFPI TKGLDEVHGYEPSIENM RRLMDVFAYNKLLTDELAEMRYQASVRPGQESFAAMFPAPRQGLDNLASPEEDIR KLPHETLILHGREDEVIPLEVLQAE LIDRAHSNVFGRCAIWTQIEHAGRFARLVNDF LHEADQPP E" 6516. .7184 /gene="atdE" 6516. .7184 /gene="atdE" /codon_start=1 /transl_table=11 /product="2-oxopent-4-dienoate hydratase" /protein_id="BAA23558.1" /db_xref="GI:2627155" /translation="MEQAKIQAFAD ELYEALVNR EAVPPLTSRADDITIEDAYHISLR MLERRQAGARIIGKKIGVTSKAVMNNMLNVHQPDFGYLTDDMVFNSGEEVNISDR LIA PRAEGEIAFILKDLTGPGVTNADVLAATECVMPCFEIVDSRIKDKWKIAIQDTIADNA SCGLFVLGDKAVSPREVDLVTGCMVVEKNGAIISTGAGAAALGSPVNCVTWLANTLGG FGIR" 7290. .8198 /gene="atdF" 7290. .8198 /gene="atdF" /codon_start=1 /transl_table=11 /product="acetolaldehyde dehydrogenase" /protein_id="BAA23559.1" /db_xref="GI:2627156" /translation="MNKKLKAFIFGPGNIGTDL LMKMORSEWIEPYVMVGVIDPESEGL KRAEMGIKTCATGVDGVLEHVIEDDIRVAFDATSAYVHAENSRKLNELGVIMIDLTP AAGVPFCVPPVNLTDHAKNLEMNVMVTCVPSHIPMVAASVRVQAVAYGEIVATVSSR SVGPQTRQNIDEFTRTAGAVEKVGGAKEGKAIIIIINPAEPP LMMRDTIHCLT ESEPD QAAITDSVHAMVKEYQYVPGYSLVNGPVEDGNKVTCTYMQVEGLGDFLPKYAGNLDIM TAAALRTAE MFAEEAANGVITLPARG" 8214. .9236 /gene="atdG" 8214. .9236 /gene="atdG" /codon_start=1 /transl_table=11 /product="4-hydroxy-2-oxovalerate aldolase" /protein_id="BAA23560.1" /db_xref="GI:2627157" /translation="MNLNGKKVTLHDMSLR DGMHAKRHQISLDENVSIATGMDEAGMP LIEVTHGDGLGGRSLNYGFPAHSDEEYLSAVVPKMKNAKISALLPLGIGTIVDHLKMAK DLGVSTIRVATHCTEADV SQOHIGMASKMGMDTVGFLMAHVMVSPENPGQAKLMVG YG ANCIYATDSAGYMLPDEV TAKIGLLRSELPSNIEVGFHGHNMGMAIANS LAARSGAS
gene	1902. .2828 /gene="atdR" 1902. .2828 /gene="atdR" /note="similar to LysR-type transcriptional activator" /codon_start=1 /transl_table=11 /protein_id="BAA23553.1" /db_xref="GI:2627150" /translation="MRCVLKHSSSYDNANWDNVKAF LALYRAKDYEA AAESLGVDGST LRRKIQSLEVVLGYSLFVKENNRWALAPGTEDVLKAAVDMESATRLFFGLTPGDKGGV IKVSIPEVLNQFSDVLI AF RKIYPEFTEDISSDAREVDL EREGDFAIRLARPI SNM NSLKI RGLFPGLGVGFSASYIDDI VDFEFGKDALRDKSDLIKTGIGFSYKPHDFV FGL LDWEOLGFGYGNIKIMCDDLESCARFCDKGAGLAILPKFLAHRYPNMSCVYDASDTLMA ELWLISRLDKSDWQVVLGNMLSAKS KEMEVS" 2888. .3226 /gene="atdS" 2888. .3226 /gene="atdS" /codon_start=1 /transl_table=11 /product="small ferredoxin-like protein" /protein_id="BAA23554.1" /db_xref="GI:2627151" /translation="MGESYQITEQCSGORFPCKAGQSVLKAMEQQGLECAPVGCRRGGG CGLCKVTVREGDYECGKMSRVHAPPEALAQGEVLACRIYPLSDLIIECRPRQSAAGLA NEYTTTKAMR" 3232. .4155 /gene="atdB" 3232. .4155 /gene="atdB" /codon_start=1 /transl_table=11 /product="catechol 2,3-dioxygenase" /protein_id="BAA23555.1" /db_xref="GI:2627152" /translation="MKKGVMRPGHVQIRVLDMDEAVKHYKDLLGLIEMDRDDQGRVYL KAWTEVDRFSVWLREADEPGMDPMGFKCLSEEVVDQLRGELAAFGCEIEDYPEDELKD CGRRVREFTAPTGHSEFIYATKQTGKGWGNHNPN EAWPRGLEGMKATRFDHCLLYGPN LDETNLNFRDVLGLDLAEQVMAPDGKRVSQFLT VSTFKAHDIAFIHHEEPKFFHASEFF LETQDVLKAADLLSMTDTSIDIGPTRHGLTHGQTIYFFDP SGNRNEVFAGGDYHYPD HEPVTDAEELGKAIFYHDRV LNERFMTVLT" 4187. .5647 /gene="atdC"	gene	gene
CDS		CDS	
gene		gene	
CDS		CDS	
gene		gene	
CDS		CDS	
gene		gene	
CDS		CDS	

RIDGSVAGLGAGAGNTPLVFCAVLDRMGVETGVLDYKIMDVAE DLVVPIMDQPIRVD
RNALTLGVAGVYSSELLFAEAEKKYGISARDILVELGRGTGGQEDMIEDLALTMA
KEKGLI"

Query Match 1.4%; Score 21; DB 1; Length 10498;
Best Local Similarity 100.0%; Pred. No. 7.8;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1201 acggaagaagagccattgag 1221
|||||
Db 5393 ACGGAGAGACAGGCCATTGAG 5413

RESULT 8
AL356308 102910 bp DNA linear PRI 14-DEC-2001
LOCUS Human DNA sequence from clone RP11-475L16 on chromosome 10,
DEFINITION complete sequence.

ACCESSION AL356308
VERSION AL356308.16 GI:17894610
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 102910)
AUTHORS Tracey,A.
TITLE Direct Submission
JOURNAL Submitted (14-DEC-2001) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

On Dec 17, 2001 this sequence version replaced gi:16972876.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >= 30);
an attempt was made to resolve all sequencing problems, such as
compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em.; EMBL; Sw.;
SWISSPROT; Tr.; TREMBL; Wp.; WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 10, constructed by the Sanger Centre Chromosome 10
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr10

RP11-475L16 is from the library RPCI-11.2 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACE3.6

IMPORTANT: This sequence is not the entire insert of clone
RP11-475L16 It may be shorter because we sequence overlapping
sections only once, except for a short overlap.
The true right end of clone RP11-475L16 is at 102910 in this
sequence. The true left end of clone RP11-557K21 is at 17539 in
this sequence. The true right end of clone RP11-146P20 is at 2000
in this sequence.

FEATURES
source
1. .102910
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-475L16"
/clone_lib="RPCI-11.2"
BASE COUNT 32050 a 20754 c 19228 g 30878 t

ORIGIN

Query Match 1.4%; Score 21; DB 9; Length 102910;
Best Local Similarity 100.0%; Pred. No. 8.7;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 830 cggcaagagatttctcaaaat 850
|||||
Db 72046 CGGCAAGGATTTCTCAAAAT 72066

RESULT 9
AC008850/c 131805 bp DNA linear PRI 25-SEP-2001
LOCUS Homo sapiens chromosome 5 clone CTD-2172I6, complete sequence.
DEFINITION AC008850
ACCESSION AC008850
VERSION AC008850.7 GI:15741477
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 131805)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 131805)
TITLE Direct Submission
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE 3 (bases 1 to 131805)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (25-SEP-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On Sep 25, 2001 this sequence version replaced gi:7711359.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov

Finishing Completed at Stanford Human Genome Center
www.sbgc.stanford.edu
Quality: Phrap Quality >=40 99.5% of Sequence;
Estimated Total Number of Errors is 0.5.
FEATURES
source
1. .131805
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTD-2172I6"
BASE COUNT 43916 a 26019 c 24296 g 37574 t
ORIGIN

Query Match 1.4%; Score 21; DB 9; Length 131805;
Best Local Similarity 100.0%; Pred. No. 8.8;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 525 ctgctccttcattctcaaacg 545
|||||
Db 88638 CTGCTCCTTCATCTTCAAAGC 88618

RESULT 10
AP003928/c 138904 bp DNA linear HTG 19-JUL-2001
LOCUS Oryza sativa chromosome 8 clone OJ1150_All, *** SEQUENCING IN
DEFINITION PROGRESS ***, in ordered pieces.
ACCESSION AP003928
VERSION AP003928.1 GI:14915711
KEYWORDS HTG; HTGS_PHASE2.
SOURCE Oryza sativa (cultivar:Nipponbare) DNA, clone:OJ1150_All.
ORGANISM Oryza sativa

REFERENCE 1 (bases 1 to 138904)
AUTHORS Sasaki,T., Matsumoto,T. and Yamamoto,K.
TITLE Oryza sativa nipponbare(GA3) genomic DNA, chromosome 8, BAC clone:OJ1150_All
JOURNAL Published Only in Database (2001) In press
REFERENCE 2 (bases 1 to 138904)
AUTHORS Sasaki,T., Matsumoto,T. and Yamamoto,K.
TITLE Direct Submission
JOURNAL Submitted (18-JUL-2001) Takuji Sasaki, National Institute of Agrobiological Resources, Rice Genome Research Program; Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:tsasakienlas.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/, Tel:81-298-38-7441, Fax:81-298-38-7468)
The nucleotide sequence of this BAC clone was generated by combining Monsanto and RGP-Japan sequencing data.
NOTE: It currently consists of 1 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have provided by the submitter. This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
source
1..138904
/organism="Oryza sativa"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/chromosome="8"
/clone="OJ1150_All"
BASE COUNT 40002 a 28323 c 28941 g 41170 t 468 others
ORIGIN

Query Match 1.4%; Score 21; DB 2; Length 138904;
Best Local Similarity 100.0%; Pred. No. 8.8;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 510 cgcctcgccgcgcgctgc 530
|||||
Db 47643 CGCCTCGCCGCGCTGCTC 47623

RESULT 11
AC008907
LOCUS Homo sapiens chromosome 5 clone CTD-2261G7, WORKING DRAFT SEQUENCE,
DEFINITION 24 unordered pieces.
ACCESSION AC008907
VERSION AC008907.5 GI:13699452
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 147344)
DOE Joint Genome Institute.
Sequencing of Human Chromosome 5
Unpublished
2 (bases 1 to 147344)
DOE Joint Genome Institute.
Direct Submission
Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Apr 20, 2001 this sequence version replaced gi:7705395.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI

Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 727924
Center clone name: CITB-H1_2261G7

Summary Statistics
Consensus quality: 122801 bases at least Q40
Consensus quality: 131739 bases at least Q30
Consensus quality: 134984 bases at least Q20
Estimated insert size: 123000; pulse field gel estimation
Estimated insert size: 145044; sum-of-contigs estimation
Quality coverage: 8.83 in Q20 bases; pulse field gel estimation
Quality coverage: 7.49 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 24 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1024: contig of 1024 bp in length
* 1025 1124: gap of unknown length
* 1125 2152: contig of 1028 bp in length
* 2153 2252: gap of unknown length
* 2253 3267: contig of 1015 bp in length
* 3268 3367: gap of unknown length
* 3368 4464: contig of 1097 bp in length
* 4465 4564: gap of unknown length
* 4565 5581: contig of 1017 bp in length
* 5582 5682: gap of unknown length
* 5683 7286: contig of 1604 bp in length
* 7287 7385: gap of unknown length
* 7386 8562: contig of 1177 bp in length
* 8563 8663: gap of unknown length
* 8664 9732: contig of 1069 bp in length
* 9733 9832: gap of unknown length
* 9833 11051: contig of 1220 bp in length
* 11052 11151: gap of unknown length
* 11152 12313: contig of 1162 bp in length
* 12314 12414: gap of unknown length
* 12415 13442: contig of 1029 bp in length
* 13443 13542: gap of unknown length
* 13543 14997: contig of 1455 bp in length
* 14998 15098: gap of unknown length
* 15099 16437: contig of 1340 bp in length
* 16438 16537: gap of unknown length
* 16538 17545: contig of 1008 bp in length
* 17546 17646: gap of unknown length
* 17647 19312: contig of 1667 bp in length
* 19313 19412: gap of unknown length
* 19413 21064: contig of 1652 bp in length
* 21065 21164: gap of unknown length
* 21165 23214: contig of 2050 bp in length
* 23215 23314: gap of unknown length
* 23315 29721: contig of 6407 bp in length
* 29722 29821: gap of unknown length
* 29822 39205: contig of 9384 bp in length
* 39206 39305: gap of unknown length
* 39306 54490: contig of 15185 bp in length
* 54491 54590: gap of unknown length
* 54591 66531: contig of 11941 bp in length
* 66532 66632: gap of unknown length
* 66633 85895: contig of 19264 bp in length
* 85896 85996: gap of unknown length
* 85997 105738: contig of 19743 bp in length
* 105739 105839: gap of unknown length
* 105839 147344: contig of 41506 bp in length.
Location/Qualifiers
1..147344
/organism="Homo sapiens"
/db_xref="taxon:9606"

/chromosome="5"
/clone="CTD-2261G7"
/clone_lib="CalTech human BAC library D"
BASE COUNT 42045 a 27306 c 28277 g 47403 t 2313 others
ORIGIN

Query Match 1.4%; Score 21; DB 2; Length 147344;
Best Local Similarity 100.0%; Pred. No. 8.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 525 ctgctctcttcattctcaagc 545
|||||
Db 140097 CTGCTCCTTCATCTCAAGC 140117

RESULT 12
AL627302 191415 bp DNA linear HTG 30-JAN-2002
LOCUS
DEFINITION Mus musculus chromosome X clone RP23-272D10, *** SEQUENCING IN
PROGRESS ***, in ordered pieces.

ACCESSION AL627302
VERSION AL627302.18 GI:18476832
KEYWORDS HTG; HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE house mouse.

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (sites)
Bates, K.

REFERENCE Direct Submission
AUTHORS Submitted (29-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
TITLE Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
JOURNAL humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Feb 1, 2002 this sequence version replaced gi:18250545.

COMMENT ----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: bm272D10

----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 99% of reads
Chemistry: Dye-primer Big Dye; 0% of reads
Consensus quality: 191248 bases at least Q40
Consensus quality: 191391 bases at least Q30
Consensus quality: 191407 bases at least Q20
Insert size: 191415; sum-of-contigs
Insert size: 182914; 6.3% error; agarose-fp
Quality coverage: 13.20x in Q20 bases; sum-of-contigs Quality
coverage: 13.82x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
source
Location/Qualifiers
1. .191415
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="X"
/clone="RP23-272D10"
/clone_lib="RPCI-23"
misc_feature 1. .191415
/note="assembly_fragment:03353
clone_end:SP6
vector_side:left"

BASE COUNT 52805 a 40506 c 40442 g 57662 t
ORIGIN

Query Match 1.4%; Score 21; DB 2; Length 191415;
Best Local Similarity 100.0%; Pred. No. 9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1125 tgtcaatcccaaacaggcag 1145
|||||
Db 105060 TGTCAATCCCAACAGCAG 105080

RESULT 13
AC008908/c 251664 bp DNA linear HTG 19-APR-2001
LOCUS
DEFINITION Homo sapiens chromosome 5 clone CTD-2265E4, WORKING DRAFT SEQUENCE,
16 ordered pieces.

ACCESSION AC008908
VERSION AC008908.5 GI:7711373
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 251664)

REFERENCE DOE Joint Genome Institute.
AUTHORS Sequencing of Human Chromosome 5
TITLE Unpublished
JOURNAL
REFERENCE 2 (bases 1 to 251664)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On May 6, 2000 this sequence version replaced gi:7709396.

COMMENT ----- Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 729409
Center clone name: CITB-H1_2265E4

Summary Statistics
Consensus quality: 232454 bases at least Q40
Consensus quality: 246311 bases at least Q30
Consensus quality: 248647 bases at least Q20
Estimated insert size: 218000; pulse field gel estimation
Estimated insert size: 250914; sum-of-contigs estimation
Quality coverage: 6.4 in Q20 bases; pulse field gel estimation
Quality coverage: 5.56 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.

* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

* 1
* 25893: contig of 25893 bp in length
* 25894 25993: gap of unknown length
* 25994 39631: contig of 13638 bp in length
* 39632 39731: gap of unknown length
* 39732 61054: contig of 21323 bp in length
* 61055 61154: gap of unknown length
* 61155 77299: contig of 16145 bp in length
* 77300 77399: gap of unknown length
* 77400 98305: contig of 20906 bp in length
* 98306 98405: gap of unknown length
* 98406 109130: contig of 10725 bp in length
* 109131 109230: gap of unknown length
* 109231 112039: contig of 2809 bp in length
* 112040 112139: gap of unknown length
* 112140 126605: contig of 14466 bp in length

* 126606 126705: gap of unknown length
* 126706 132688: contig of 5983 bp in length
* 132689 132788: gap of unknown length
* 132789 145777: contig of 12989 bp in length
* 145778 145877: gap of unknown length
* 145878 150206: contig of 4329 bp in length
* 150207 150306: gap of unknown length
* 150307 180051: contig of 29745 bp in length
* 180052 180151: gap of unknown length
* 180152 208645: contig of 28494 bp in length
* 208646 208745: gap of unknown length
* 208746 217750: contig of 9005 bp in length
* 217751 217850: gap of unknown length
* 217851 237954: contig of 20104 bp in length
* 237955 238054: gap of unknown length
* 238055 251664: contig of 13610 bp in length.

FEATURES
source
1..251664
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTD-226584"
/clone_lib="CalTech human BAC library D"
BASE COUNT 80411 a 47554 c 46337 g 75858 t 1504 others
ORIGIN

Query Match 1.4%; Score 21; DB 2; Length 251664;
Best Local Similarity 100.0%; Pred. No. 9.1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 525 ctgctccttcattcttcaagc 545
|||||
Db 186250 CTGCTCCTTCATCTTCAAAGC 186230

RESULT 14
FVNS5GAI
LOCUS FVNS5GAI 481 bp ss-RNA linear VRL 07-JUN-1996
DEFINITION Sepik virus nonstructural protein (NS5) gene, 3' end of cds.
ACCESSION L48981
VERSION L48981.1 GI:1066550
KEYWORDS NS5 gene; nonstructural protein.
SOURCE Sepik virus (strain MK7148) cDNA to genomic RNA.
ORGANISM Sepik virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Flavivirus; mosquito-borne viruses.
1 (bases 1 to 481)
Poidinger,M., Hall,R.A. and Mackenzie,J.S.
Molecular characterization of the Japanese encephalitis serocomplex
of the flavivirus genus
Virology 218 (2), 417-421 (1996)
JOURNAL 96193756
MEDLINE
FEATURES
source
1..481
/organism="Sepik virus"
/strain="MK7148"
/db_xref="taxon:44026"
1..251
/gene="NS5"
<1..251
/gene="NS5"
/codon_start=3
/product="nonstructural protein"
/protein_id="AAB02061.1"
/db_xref="GI:1066551"
/translation="LXLSGELAXTRPCLWTKPGISXGGTTQRASWARDVKDAGHRIRK
LVGNEKFYDYMSTMGRYRELGEEGPGELLVDRLTQHP"
BASE COUNT 133 a 122 c 127 g 87 t 12 others
ORIGIN

Query Match 1.4%; Score 20; DB 14; Length 481;

Best Local Similarity 100.0%; Pred. No. 25;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1390 agagagctagcggaagaagg 1409
|||||
Db 186 ACAGAGCTAGCGGAAGAAGG 205

RESULT 15
AB009654
LOCUS AB009654 2001 bp DNA linear BCT 11-MAY-2000
DEFINITION Alteromonas sp. DNA for aldehyde dehydrogenase, complete cds.
ACCESSION AB009654
VERSION AB009654.1 GI:2696799
KEYWORDS aldehyde dehydrogenase.
SOURCE Alteromonas sp. (strain:KE10) DNA.
ORGANISM Alteromonas sp.
Bacteria; Proteobacteria; gamma subdivision; Alteromonadaceae;
Alteromonas.
1 (sites)
AUTHORS Maeda,T., Yoshinaga,I., Shiba,T., Murakami,M., Wada,A. and Ishida,Y.
TITLE Cloning and sequencing of the gene encoding an aldehyde dehydrogenase that is induced by growing Alteromonas sp. Strain KE10 in a low concentration of organic nutrients
Appl. Environ. Microbiol. 66 (5), 1883-1889 (2000)
JOURNAL 20250629
MEDLINE
REFERENCE 2 (bases 1 to 2001)
AUTHORS Maeda,T., Yoshinaga,I., Shiba,T., Murakami,M. and Ishida,Y.
TITLE Direct Submission
JOURNAL Submitted (09-DEC-1997) Toshimichi Maeda, National Fisheries University, Department of Food Science and Technology; Nagata-honmachi 2-7-1, Shimonoseki, Yamaguchi 759-6595, Japan (E-mail:toshima@fish-u.ac.jp, Tel:81-832-86-5111, Fax:81-832-86-7434)

FEATURES
source
1..2001
/organism="Alteromonas sp."
/strain="KE10"
/db_xref="taxon:232"
218..223
/note="putative"
232..1749
/gene="olga"
232..1749
/gene="olga"
/codon_start=1
/transl_table=11
/product="aldehyde dehydrogenase"
/protein_id="BAA24014.1"
/db_xref="GI:2696800"
/translation="MIYAKPGSEGSVVSFKESYCNFINGEWAPVKGEYFTATSPVDG
SDIAQVPRSTSEDELDAHAARKGAEKSVAEKSNVLLKIADRIENLELLAVAE
TWDNGKAVRETNLNADIPLCADHFRYYAGCIRSQGTIGEIDKNTVAYHFHEPYGVVGG
IIPWNPPLMAAWKLAPALVTGNVVLKPAEQTPVSILLFAELIQDLPPGVLNIVNG
FGAAGQALATSTRIAKIATGTPVGGHILRCAENLIPSTVELGKSPNLFADIM
DHEPEFIDKCAEGLVLFNQGEVCTCPSRAIIQEDIDYDAFMKVLEIKIKRGNPL
DTETMVGAQASNEQFEKILKYIDIGKSEGAQVVTGGEKEQLPGFDDGYIIQPTILKGE
NHRVFOEEIFGPVIAVQTFKTEEEAIEIANNTSFGLGAGVWSRDANVAFRVGRAIEA
GRVWTNCYHLYPAAHAFGGYKKSIGRETHKMLDHYQOTKNLLVSYDINPLGFF"
BASE COUNT 551 a 461 c 487 g 502 t
ORIGIN

Query Match 1.4%; Score 20; DB 1; Length 2001;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1201 acggaagaagagccattga 1220
|||||
Db 1471 ACGGAAGAAGAGGCCATTGA 1490

Search completed: April 27, 2002, 05:34:37
Job time: 17314 sec

Db 191 RVVVMGTGASGIQVIOEAAVAEHLTVFQRTPNLALPMRQORLSADDNDRYENIEDRFQ 250
QY 244 ASKRSRIGFPVQAPSVGIFEVSPQREAYFEELWERGAFNFLACQYREVVMVDKKANRLVY 303
Db 251 IRONSFAGDFYFIPQNAADTPEDERTAIYEKMWDEGGFPLWLNFGQLLTDEAAHHTFY 310
QY 304 DFWAKTRSRIVNPAKRDLMAPLEPPYWFGTKRSPLESDYEMLDKPSVEIVNLEQSPIV 363
Db 311 NEWRSKVHDRVKDPKTAEMLAPATPPHPFGVKRPSLEQNYFDVYNQDNVDLIDSNATPIT 370
QY 364 AVTKTGVLSDGSKRECDTIVLATGDFDSTGSLTHMGLKNKHGVDLKEVWKDGIISTYMGV 423
Db 371 RVLPNGVETPDGVV-ECDVLVLATGFDNNSGGINAIDIK-AGGQLLRDKWATGVDTYMG 428
QY 424 FSHGFPNFAFFVATAQAPTIVLSNGPTIETQVDLIADTIKLEAEHATSVEATKSAQEAWS 483
Db 429 STHGFPNLMFLYGPQSPGFCNGTDEGGAPGDMVADFLWLKDNKGISRFESTEEREWR 488
QY 484 IMIAKMNHEHTLFPLTDSWWTGGNIPGKATRALTF 517
Db 489 AHVDDIFVNSLFPPKAKSWYGANVPGKPAQMLNY 522

RESULT 6
Q9R2F5 PRELIMINARY; PRT; 543 AA.
ID Q9R2F5
AC Q9R2F5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CYCLOHEXANONE 1,2-MONOXYGENASE.
GN CHNB.
OS Acinetobacter sp., and
OS Acinetobacter sp. NCIMB9871.
OC Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
OC Acinetobacter.
OX NCBI_TaxID=472, 93373;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=Acinetobacter sp.; STRAIN=NCIB 9871;
RX MEDLINE=20011264; PubMed=10543838;
RA Iwaki H., Hasegawa Y., Teraoka M., Tokuyama T., Bergeron H.,
RA Lau P.C.K.;
RT "Identification of a transcriptional activator (ChnR) and a 6-oxohexanoate dehydrogenase (ChnE) in the cyclohexanol catabolic pathway in Acinetobacter sp. strain NCIMB 9871 and localization of the genes that encode them.";
RL Appl. Environ. Microbiol. 65:5158-5162(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=Acinetobacter sp.; STRAIN=NCIB 9871;
RX MEDLINE=88115180; PubMed=3338974;
RA Chen Y.-C.J., Peoples O.P., Walsh C.T.;
RT "Acinetobacter cyclohexanone monooxygenase: Gene cloning and sequence determination.";
RL J. Bacteriol. 170:781-789(1988).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=Acinetobacter sp. NCIMB9871; STRAIN=NCIMB 9871;
RA Iwaki H., Hasegawa Y., Teraoka M., Tokuyama T., Lau P.C.;
RT "Identification and Characterization of Cyclohexanol Metabolic Genes from Acinetobacter sp. NCIMB 9871.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
CC -1- COFACTOR: FAD (BY SIMILARITY).
DR EMBL: AB006902; BAA86293.1; -.
DR EMBL: AB026668; BAB61738.1; -.
DR InterPro: IPR001327; FAD_pyr_redox.
DR InterPro: IPR000960; FMO.
DR InterPro: IPR001100; pyr_redox.
DR Pfam: PF00743; FMO-like; 1.
DR PRINTS; PR00368; FADPNR.
DR PRINTS; PR00370; FMOXYGENASE.
DR PRINTS; PR00411; PNDRDTASEI.

KW FAD; Flavoprotein; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 543 AA; 60937 MW; EC9AEABE49A270AF CRC64;
Query Match 27.8%; Score 806; DB 2; Length 543;
Best Local Similarity 33.2%; Pred. No. 3.6e-58;
Matches 179; Conservative 100; Mismatches 230; Indels 30; Gaps 8;
QY 16 DAIIVGAGLSGISAVYKLR-KLRINAKIFEGAPDFGGVWHNRYPGARVDSETPFY---- 70
Db 8 DAIVIGGGGGLYAVKKLRDELEKLVQAFDKATDVAGTWYWNRYPCALTDTEHLYCYSW 67
QY 71 ---QLNIPEVWKDWTWSCRYPDQKELLSYVHHCDKIRGLRKDVYFGAEVVDARYARDLGT 127
Db 68 DKELLQSLKIKK-----KYVOGPDVRKYLQQAEEKHDLKKSYPNTAVQSAHYNEADAL 121
QY 128 WTVKTSAGHVATAKYLIATGLLHRRKHTPALPLADFNKGVIHSSAWHEDFDAEGORVAV 187
Db 122 WEVTTEYGDKYTARFLITAGLLSAPNLPNIKGINQFKGELHHTSRWPDVVSFEGKRVGV 181
QY 188 IGAGATSIQIVQELAKKADQVTMFMRHPSYCLPMRQRTMDRNEQTAKKAYYPTLFEASRK 247
Db 182 IGTGSTGVQVITAVAPLAKHLITVEQSAQYSVPIGNOPLSEEDVKKIKDNYDKIMDGVVN 241
QY 248 SRIGFPVQAPSVGIFEVSPQREAYFEELWER-GAFNFLACQYREVVMVDKKANRLVYDFW 306
Db 242 SALAFGLNESTVPAMSVSAEERKAVFEKAWOTGGGFRFMETFGDIATNMEANIEAQNFI 301
QY 307 AKKTRSRIVNPAKRDLMAPLEPPYWFGTKRSPLESDYEMLDKPSVEIVNLEQSPIVAVT 366
Db 302 KGKIAEIVKDPPIAQKLMPOD----LYAKRPLCDSGYYNTFNDRNVRLVEDVKANPIVEIT 357
QY 367 KTGVLSDGSKRECDTIVLATGDFDSTGSLTHMGLKNKHGVDLKEVWKDGIISTYMGVFSH 426
Db 358 ENGKLENGDFVELDMLICATGDAVDGNYVRMDIOCKNGLAMKDYWKKEGPPSSYMGVTVN 417
QY 427 GFPNAFFVATAQAPTIVLSNGPTIETQVDLIADTIKLEAEHATSVEATKSAQEAWSIMI 486
Db 418 NYPNMFVGLPNGP--FTNLPSPSIESQVEMISDTIOYTVENNVSIEATKEAEQWTQTC 475
QY 487 AKMNEHTLFPLTDSWWTGGNIPGKATRALTFIGGIALYEIQICEQKVN-----WDGFDV 540
Db 476 ANIAEMTLFPKQSWIFGANIPGKNITVYFYLGLKEY----RSALANCKNHAYEGFDI 530
RESULT 7
Q9F7E4
ID Q9F7E4 PRELIMINARY; PRT; 543 AA.
AC Q9F7E4;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CYCLOHEXANONE MONOOXYGENASE.
GN CHNB.
OS Acinetobacter sp. SE19.
OC Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
OC Acinetobacter.
OX NCBI_TaxID=135835;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SE19;
RX MEDLINE=20398154; PubMed=10940013;
RA Cheng Q., Thomas S.M., Kostichka K., Valentine J.R., Nagarajan V.;
RT "Genetic analysis of a gene cluster for cyclohexanol oxidation in acinetobacter sp. strain SE19 by in vitro transposition.";
RL J. Bacteriol. 182:4744-4751(2000).
CC -1- COFACTOR: FAD (BY SIMILARITY).
DR EMBL: AF282240; AAG10021.1; -.
DR InterPro: IPR001327; FAD_pyr_redox.
DR InterPro: IPR000960; FMO.
DR InterPro: IPR001100; pyr_redox.
DR PRINTS; PR00368; FADPNR.
DR PRINTS; PR00370; FMOXYGENASE.


```
RT "Complete genome sequence of Caulobacter crescentus.";
RA Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
CC -!- COFACTOR: FAD (BY SIMILARITY).
DR EMBL; AE005856; AAK23793.1; -.
DR TIGR; CC1818; -.
DR InterPro; IPR001327; FAD_pyr_redox.
DR InterPro; IPR000960; FMO.
DR InterPro; IPR001100; pyr_redox.
DR Pfam; PF00743; FMO-like; 1.
DR PRINTS; PR00368; FADPNR.
DR PRINTS; PR00411; PNDRTASEI.
KW Complete proteome; FAD; Flavoprotein; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 543 AA; 59597 MW; 946E00ABD5AC877B CRC64;

Query Match 27.1%; Score 786; DB 16; Length 543;
Best Local Similarity 33.3%; Pred. NO. 1.6e-56;
Matches 172; Conservative 94; Mismatches 240; Indels 10; Gaps 6;

QY 17 AIIVGAGLSGISAVYKLRKRLNAKIFEGAPDEGGVWNNRYPGARVDSETPFYQLNIPE 76
   ||:|||| |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 18 AIIVGAGFGGMYAIHRLRQEGSLVGLGAGDVGCTWYNNRYPGARCDVTSIDISYAFSQ 77

QY 77 -VWKDWTWSCRYPDQKELLSYVHHC-KIRGLKQDVYFGAEVVDARYARDLGTWTVKTSAG 135
   ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 78 AIEQEWTSSEVFAGSEILAYANFVADRLDLRSAYQFNTRVQTIAFDDARRVWTLTDDQ 137

QY 136 HVATAKYLLILATGLLHRKHTPALPGLADFNKGVIHSSAW-HEDFQAEQORVAVIGAGATS 194
   | ||:|||| |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 138 QRFEADNLIMATGSLPKPVDIPGADRFKGLYRSCKRPHHAFVDFSGKRVGVVGTGSSG 197

QY 195 IQIVQELAKKADQVTFMFMRRPSYCLPMRQRTMDRNEQTAWKAYPTLFEASRKSRI 254
   ||||:|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 198 IQIVPVVAETCQSLTVFQRTSPSTLPMRNKKVSEDYAAQLKANMAGLRAAAARTTFTGG 257

QY 255 QAPSVGIFEVSPEQREAYFEELWERGAFNFLACQYREVVMVDKKANRLVYDFWAKKTRRSR 314
   ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 258 PLSTRPLFSIPVEERNELMEEAWAKALQFLGL-FSDLLTNAANAIEIVADFVRDKISQTV 316

QY 315 VNPAKRDLMAPLEPPYFVGTKRSPLESDDYEMLDKPSVEIVNLEQSPIVAVTKTGVLSD 374
   ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 317 KDPVTAEALKPRGYPIY--ARRPCLDTQYYETYNLPHVSLYDCFEETISFTETGLRT-- 372

QY 375 GSKRE--CDTIVLATGDFDSTGSLTHMGLKNKHGVDLKEVWKDGISTYMGVFSHGFPNAF 432
   ||||| |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 373 -SQREIALDVIIMATGYDALTGPMALDITGROGLKLEAWRDGPRSYLALLMHGFPNLF 431

QY 433 FVATAQAPTVLSNGPTIIEETQVLDIADTIKLEAEHATSVEATKSAQEAWSIMIAKMNEH 492
   ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 432 IIGGPNGPSILANVRLDELNVVDVYDCILHMQRGLATVETAEATQTRNSQTVGEIADR 491

QY 493 TLEPPLTDSWWTGGNIPGKATRALTFIGGIALYEIC 528
   ||:|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 492 TLVVKADTWYTGNTIEGKPRGFFVFAGGLNRYAEIC 527

RESULT 10
Q9I218 PRELIMINARY; PRT; 491 AA.
AC Q9I218;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE PROBABLE FLAVIN-BINDING MONOOXYGENASE.
GN PA2097.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
```

```
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Coltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saler M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
CC -!- COFACTOR: FAD (BY SIMILARITY).
DR EMBL; AE004637; AAG05485.1; -.
DR InterPro; IPR000759; Adrndx_reductase.
DR InterPro; IPR001327; FAD_pyr_redox.
DR InterPro; IPR000960; FMO.
DR InterPro; IPR000205; NAD_binding.
DR InterPro; IPR00103; pyridine_redox_2.
DR InterPro; IPR001100; pyr_redox.
DR Pfam; PF00743; FMO-like; 1.
DR PRINTS; PR00419; ADXRDTASE.
DR PRINTS; PR00368; FADPNR.
DR PRINTS; PR00411; PNDRTASEI.
DR PRINTS; PR00469; PNDRTASEII.
KW Complete proteome; FAD; Flavoprotein; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 491 AA; 56082 MW; 1CB341148CD237DA CRC64;

Query Match 19.0%; Score 550.5; DB 16; Length 491;
Best Local Similarity 28.6%; Pred. NO. 4.6e-37;
Matches 154; Conservative 84; Mismatches 210; Indels 91; Gaps 17;

QY 18 IIIVGAGLSGISAVYKLRKRLNAKIFEGAPDFGCVWNNRYPGARVDSETPFYQLNI-P 75
   :|||:|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 14 LVVGAGFAGLGLAIRLQAGIDDDLVLEKAADVGGCWNRNRYPGACDVPSHLYSFSFEP 73

QY 76 EVWKDWTWSCRYPDQKELLSYVHHC-DKIRGLKRDVYFGAEVVDARYARDLGTWTVKTS 134
   :||:|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 74 KA---DWSRKFAPQAEILDLRHCDKYR-LREKIRFHCEVEEARFDAGSGEQVRCAD 128

QY 135 GHVATAKYLLILATGLLHRKHTPALPGLADFNKGVIHSSAWHEDFQAEQORVAVIGAGATS 194
   | |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 129 GQTLRARALVCATGQLSRPLLPRLPGLERFQGPAPAFHSANWDAEVELAGKRVAVIGTGA 186

QY 195 IQIVQELAKKADQVTFMFMRRPSYCLPMRQRTMDRNEQTAWKAYPTLFEASRKSRI 254
   || |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 189 IQFVPRIAPQVQRLSLFQRSAPYVIAKPDRTYADWERRL-KARWPWLQRLDR----- 239

QY 255 QAPSVGIFEVSPEQREAYFEELWERGAFNFLACQYREVVMVDKKANRLVYDFWAKKTRRSR 314
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 240 ----GLKYLHHSRMLAF-----ATFPAL-----MKVMRLSFH---RHLHRQI 275

QY 315 VNPAKRDLMAPLEPPYFVGTKRSPLESDDYEMLDKPSVEIVNLEQSPIVAVTKTGVLSD 374
   :| |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 276 ADP--QLRARLVDPYPLGCKRILISNDYYPALARSNVELVD--TGIREVTEDAVVGRD 329

QY 375 GSKRECDTIVLATGDFDSTGSLTHMGLKNKHGVDLKEVWKDGISTYMGVFSHGFPNAFFV 434
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 330 GRRHEVDALIFGTGF-AATEPLAPMRILGLDGRDLRQAWADGAAYKGISVSGFPNLFIL 388

QY 435 ATAQAAPTIVLSNGPT-----IIEETQVLDIADTIKLEAEHATSVEATKSAQEAWSI 484
   || |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 389 -----YGPNTNLGHNSIVYMLESQFPYVLGCLRLQLOEQGLRDLVKPEVQRRRNL 438

QY 485 MIAKMNEHTLFPPL-TDSWWTGGNIPGKATRALTFIGGIALYEICQEKVANWGDGFDVLH 542
   : |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 439 EVQQGLRHTVWERGCDSW-----YKTAACKNTNNWPGYTFVY 475

RESULT 11
Q9RKB5 PRELIMINARY; PRT; 519 AA.
ID Q9RKB5
AC Q9RKB5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
```



```
QY 247 KSRIGFPVQAPSQVGEIEVSPEQREAYFEELWERGAFNFLACQYREVMVDKKANRLVYDFW 306
Db 240 EAT--FPIAAHYFAVEPLAKHMSA-----G 263

QY 307 AKTRSRIVNPAKRDLMAPLEPPYWFQTKRSPLESDYYEMLDKPSVEIVNLEQSPIVAVT 366
Db 264 RRYLRQVHDPVRE--QLTPRYAVGCKRPGFHNLYSTFNDRNRLVT--EPIDKIT 317

QY 367 KTGVLSDGSKREDTIVLATGFSF--TGSILTHMGLKKNKHGVDLKEVWKD-GISTYMGVF 424
Db 318 PTAVATTDGASHEIDVLVLATGFKVLDTSIPTYAVTGTGGASLSRFWDEHRLQAYEGVS 377

QY 425 SHGFPNAFFVATAQAPTIVLSNGPTIETQVDLIADTIAKLEAEHATSVEATKSAQEAWSI 484
Db 378 VPGYPN-FFTVEGPGYGVGSSYFALIEQAHRIIRCLKRARRTGATRIEVTTEANARYFA 436

QY 485 MIAKMEHTLF-----PLTDSW--TGGNIPGKAT 512
Db 437 EVMRRRHQVFWQDSCLANSYYFDKNGDVPLRPT 471

RESULT 13
O53294 PRELIMINARY; PRT; 524 AA.
AC O53294;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DE MONOXYGENASE.
GN RV3049C OR MT012.64C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
DR EMBL; AL021287; CAA16134.1; -.
DR TubercuList; Rv3049c; -.
DR InterPro; IPR000960; FMO.
DR Pfam; PF00743; FMO-like; 1.
KW Complete proteome.
SQ SEQUENCE 524 AA; 58776 MW; BABAF86F4DD1AB54 CRC64;
```

```
Query Match 17.1%; Score 497.5; DB 16; Length 524;
Best Local Similarity 28.7%; Pred. No. 1.3e-32;
Matches 142; Conservative 78; Mismatches 218; Indels 57; Gaps 13;

QY 17 AIIVGAGLSGSAVYKLRKRLNNAKIFEGAPDFGGVWHNNRYPGARVDSETPFYQLNIPE 76
Db 25 AVIICTGFGSLGMAIALQKQGVDFVILEKADDDVGGTWRDNTYPCACDIPSHLYSFSF-E 83

QY 77 VKDWTWSCRYPDQKELLSYVHHCDKIRGLRKDVYFGAEVVDARYARDLGTWTVKTSA 136
Db 84 PKADWKHLFSYND--EILGYLKGVTQYGLRRIYEFNSLVDRGYWDDDCRWHVFTADGR 141

QY 137 VATAKYLILATGLLHRKHTPALPGLADFNKVIHSSAWHEDFAEGQORVAVIGAGATSIQ 196
Db 142 EYVAQFLISGAGALHIPSFPFIEIAGRDEFAFPFSAQWHDHSIDLTKRVAIVCTGASATQ 201
```

```
QY 197 IVOELAKKADQVMTMEMRRPSYCLPMRQRTMDRNEQTAWKAYYPTLFEASRKSRIQFPVQA 256
Db 202 IVPEIVQVAELQLYQRTPPVWVVRTNEELPVSRLRALRT-VPGLRALLR----- 250

QY 257 PSVGIFEVSPEQREAYFEELWERGAFNFLACQYREVMVDKKANRL-VYDFNAKKTSSRIV 315
Db 251 --LGIY-----WAQELAY-----GMTKRPNTLKIIIEAVAKYNIRRSV 286

QY 316 NPAKRDLMAPLEPPYWFQTKRSPLESDYYEMLDKPSVEIVNLEQSPIVAVTKGVLLSDG 375
Db 287 K--DRELRRKLTTRYIGCKRILNSSYYPAVADPKTELIT--DRIDRITHDGIIVTADG 341

QY 376 SK---RECDTIVLATGFSFTGSLTHMGLKKNKHGVDLKEVW-KDGISTYMGVFSHGFPN 430
Db 342 TGREVFREADVIVYATGF-HVTDSYTYVQIKGRHGEDLVDRWNREGIAGHARGITVANMPN 400

QY 431 APFVATAQAPTIVLSNGPTIETQVDLIADTIAKLEAEHATSVEATKSAQEAWSIMIKNM 490
Db 401 LFFLLGPNTGLGHNSVVFMIESIHYVADAIKACDRMGVQALAPTREAQD-----RFN 453

QY 491 EHTLFPPLTDSWWTGG 505
Db 454 QELQRRLAGSVVNSG 468

RESULT 14
O87636 PRELIMINARY; PRT; 437 AA.
AC O87636;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CYCLOHEXANONE MONOOXYGENASE HOMOLOG (FRAGMENT).
OS Pseudomonas fluorescens.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=294;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DSM50106;
RX MEDLINE=99124577; PubMed=9925571;
RA Khaimeyer V., Fischer I., Bornscheuer U.T., Altenbuchner J.;
RT "Screening, nucleotide sequence, and biochemical characterization of
RT an esterase from Pseudomonas fluorescens with high activity towards
RT lactones.";
RL Appl. Environ. Microbiol. 65:477-482(1999).
DR EMBL; AF090329; AAC36351.1; -.
DR InterPro; IPR000960; FMO.
DR InterPro; IPR000103; Pyridine_redox_2.
DR Pfam; PF00743; FMO-like; 1.
DR PRINTS; PR00469; PNDRDTASEII.
KW Monooxygenase.
FT NON_TER 437 437
SQ SEQUENCE 437 AA; 48863 MW; 11E72DF34C0CD13C CRC64;
```

```
Query Match 16.8%; Score 486.5; DB 2; Length 437;
Best Local Similarity 27.6%; Pred. No. 7.7e-32;
Matches 137; Conservative 79; Mismatches 204; Indels 77; Gaps 14;

QY 1 MSATNSNRGDCSVACDAIIVGAGLSGISAVYKLRKRLN-NAKIFEGAPDFGGVWHNNRYP 59
Db 1 MNAKSDS-----IDAIIGSGFAGLCMAIKLEAGFTDLFVAEQADTLGGTWRDNHYP 53

QY 60 GARVDSETPFYQLNI---PEVWKDWTWSCRYPDQKELLSYVHHCDKIRGLRKDVYFGAEV 116
Db 54 GCACDVQSHVYSFSPAPND-----WTRQFAPAEIRAYLEDCAVRFGAPYLRFGMGL 107

QY 117 VDARYARDLGTWTVKTSAHVATAKYLILATGLLHRKHTPALPGLADFNKVIHSSAWHE 176
Db 108 KRAVFDEQLRWQLSFSDGRHVSARVLVSCMGALARPALPEIPGLETFKGRFRHSQQWDH 167

QY 177 DFDAEQORVAVIGAGATSIQIVQELAKKADQVTFMRRRPSYCLPMRQRTMDRNEQTAWKA 236
```

Db 168 AYALKGKRVAVIGTGASAIQFVQIAPOVAHLDLFQRTPPWIMPKPDRGISAFER--W-- 223
QY 237 YYPTLFEASRKSRIGFPVQAPSVGIFEVSPEQREAYFEELWERGAFNFLACQYREV---- 292
Db 224 -----LFR-----HLPVTQRLV-----RCAF-YWALEGRVLGFAL 252
QY 293 -----MVDDKANRLVYDFWAKKTRSRIVNPAKRDLMAPLEPPYWFGTKRSPLESDDYEM 346
Db 253 HPQLMKMVQKVALR-----HLRKQVPRPS---LRKALTPDYTGCKRVLISNDYYP 301
QY 347 LDKPSVEIVNLEQSPIVAVTKTGVLLSDGSKRECDTIVLATGFDSTGSLTHMGLKNKHG 406
Db 302 LSRSNVEVVT---DKILRIEADGVITADGKIKHPADCLIFGTGFQA-TDPLPRDCIIGRDG 357
QY 407 VDLKEVWKDGISTYMGVFSHGFPNAPFFVATAQAAPTIVLSNGPTTIIETQVDLIADTIAKLEA 466
Db 358 VDLMDTWRDGAHAYKGTTPGYPNLFILIGPNTGLGHNSMILMIEAQVYILDALROMOR 417
QY 467 EHATSVEATKSAQEAWS 483
Db 418 HRIATVDVKPMVEQAYN 434

RESULT 15

Q938F6 PRELIMINARY; PRT; 603 AA.
AC Q938F6;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CYCLODODECANONE MONOOXYGENASE.
GN CDDA.
OS Rhodococcus ruber.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Nocardiaceae; Rhodococcus.
OX NCBI_TaxID=1830;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SC1;
RX MEDLINE=21475793; PubMed=11591693;
RA Kostichka K., Thomas S.M., Gibson K.J., Nagarajan V., Cheng Q.;
RT "Cloning and Characterization of a Gene Cluster for Cyclododecanone
RT Oxidation in Rhodococcus ruber SC1.";
RL J. Bacteriol. 183:6478-6486(2001).
DR EMBL; AY052630; AAL14233.1; -.
KW Monooxygenase.
SQ SEQUENCE 603 AA; 67536 MW; 29D169ACFE0BD3B2 CRC64;

Query Match 16.6%; Score 481.5; DB 2; Length 603;
Best Local Similarity 26.4%; Pred. NO. 3.3e-31;
Matches 153; Conservative 85; Mismatches 221; Indels 121; Gaps 19;

QY 20 VGAGLSGISAVYKLRKRL-NAKIFEGAPDFGGVYWHNRYPGARVDSSTPFYQLNIPEVW 78
Db 62 IGGGFSGLVTAARLRRESGVESVRIIDKAGDEGGVYWNRYPGAMCDTAAWVYMPLEET- 120
QY 79 KDWTSWCRYPDQKELLSVHHCDKIRLRKDVY---FGAEVVDARYARDLGTWTVKTS 134
Db 121 -GYMTEKYANGPEIL---EHCQRI-GKHYDLYDDALFTEVTDLVWQEHQDQWRISTNR 175
QY 135 GHVATAKYLILATGLLHRKHTPALPLGLADFNKVIHSSAWHEDF---DAEG-----Q 183
Db 176 GDHFTAQFVGMGTGPLHVAQLPGIPGIESFRGKSFHTSRWDYDYGTDGALGAPMDKLADK 235
QY 184 RVAVIGAGATSIQIVQELAKKADQVTFMRRPSYC-----LPMRQRTMDRNEQTAWKAYY 238
Db 236 RVAVIGTGATAVQCVPPELAKYCRELYVVVQRTPSAVDERGNHPIDEKWFAQIATPGWQKRW 295
QY 239 PTLFEASRKSRIGFPVQ-----APSVGIFEVSPEQREAYFEE 275
Db 296 LDSFTAIWDGVLTDPSLAIEHEDLVQDGWTALGQRMRAAVGSVPLEQYSPENVQRALEE 355

QY 276 LWERGAFNFLACQYREVVMVDKKANRLVYDFWAKKTRSRIVNPAKRDLMAPLEPPYWFG-- 333
Db 356 -----ADDEQMERI-----RAR-VDEIVTDPATAAQLKAWFROM 388
QY 334 TKRSPLESDDYEMLDKPSVEIVNLEQSPIVAVTKTGVLLSDGSKRECDTIVLATGFDSE- 392
Db 389 CKRPCFHDDYLPAFNRPNTHLVDTGGKGVVERITENGVVVA-GVEYEVDCIVYASGFEFLG 447
QY 393 TGSLSLTHMGL--KNKHGVDLKEVWKDGISTYMGVFSHGFPNAPFFVATAQAAPTIVLSNGPTII 450
Db 448 TGYTDRAGFDPTGRDGVKLSHWAQQTRTLHGMHTYGFNPFLVQLMQGAALGSNIPHNF 507
QY 451 ETQVDLIADTIAKLEAEHATSVEATKSAQEAWSIMI-----A 487
Db 508 VEAARVVAIVDHLVSLSTGTSSVETTKAEAEQAWVQLLLDHGRPLGNPECTPGYNNNEGKPA 567
QY 488 KMNEH-----TLFPLTDSWWTGGNIPGKATRALTF 517
Db 568 ELKDRLNVGYPAGSAAFFRMDHDLAAGSFDG-----LTF 602

Search completed: April 27, 2002, 07:57:41
Job time: 697 sec


```
QY 128 WT VKTSAGHVATAKYLI L ATGLLH R KHTPALPGLADFN GKVIHSSAWHEDFDAEQRVAV 187
D 121 WEVTEYGD KYTARFLITAGLLSAPNLPNIKGINQFKGELHHTSRWPDVDFEGKRVGV 180
QY 188 IGAGATSIQIVQELAKKADQVTFMWRPSCYCLPMRQRTMDRNEQTAWKAYYPTLFEASRK 247
D 181 IGTGSTGVQVITAVAPLAKHLTVFQRSAQYSVPIGNDPLSEEDVKKIKNDYKSLGWC MN 240
QY 248 SRIGFPPVQAPSVGIFEVSPEQREAYFEELWER-GAFNFLACQYREVVMVDKKANRLVYDFW 306
D 241 SALAFALNESTVPAMSVSAERKAVFEKAWQTGGGFREMPFETFGDIATNMEANIEAQNEFI 300
QY 307 AKKTRSRIVNPAKRDLMAPLEPPYWFGTGRSPLESDDYVEMLDKPSVEIVNLEQSPIVAVT 366
D 301 KGKIAEIVKDP AIAQKLM PQD---LYAKRPLCDSGYNTFN RDNVRLEDVKANPIVEIT 356
QY 367 KTGVLSDGSKRECDTIIVLATGFDSTFTGSLTHMGLKNKHGVDLKEVWKDGISTYMGVFESH 426
D 357 ENGVKLENGDFVELDMLCATGFDVADGNYVVRMDIQGNGLAMKDYWKEGPSSYMGVTVN 416
QY 427 GPNPAFFVATAQAAPTIVLSNGPTIETQVDLIADTIAKLEAEHATSVEATKSAQEAWSIM I 486
D 417 NYPNMFVLPNGP--FTNLPSSIESQVEVWISDTIQYTVENNVSIEATKEAEQWTQTC 474
QY 487 AKMNEHTLFPLTDSWNTGGINPGKATRALTFICGGTIALYEQICQEKVAN--WDGFDV 540
D 475 ANIAEMTLFPKQAQSWIEGANIPGKKNTVYFYLGLKEY-RTCASNCKNHAYEGFDI 529

RESULT 2
Y892_MYCTU
ID Y892_MYCTU STANDARD; PRT; 495 AA.
AC Q10532;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable monooxygenase Rv0892 (EC 1.14.13.-).
GN Rv0892 OR MT0916 OR MTCY31.20.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- COFACTOR: FAD (BY SIMILARITY).
CC -!- SIMILARITY: TO ACINETOBACTER SP. CYCLOHEXANONE MONOOXYGENASE,
CC A.NIDULANS STCW AND TO MAMMALIAN DIMETHYLANILINE MONOOXYGENASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; 273101; CAA97398.1; -.
DR EMBL; AE006978; AAK45162.1; -.
DR TIGR; MT0916; -.
DR TubercuList; Rv0892; -.
DR InterPro; IPR000960; FMO.
DR InterPro; IPR000205; NAD_binding.
DR Pfam; PF00743; FMO-like; 1.
KW Hypothetical protein; Oxidoreductase; Monooxygenase; NADP;
KW Flavoprotein; FAD; Complete proteome.
FT NP_BIND 170 202 NADP (BY SIMILARITY).
SQ SEQUENCE 495 AA; 55039 MW; 08CB55F48674DB44 CRC64;

Query Match 17.2%; Score 500.5; DB 1; Length 495;
Best Local Similarity 28.8%; Pred. NO. 1.5e-31;
Matches 143; Conservative 77; Mismatches 217; Indels 59; Gaps 12;

QY 15 CD AI-IVGAGLSGISAVYKLRKRL-NAKIFEGAPDFCGVWHNRYPCARVDSETPFYQL 72
D 5 CPTVAVVGAGSMGCMVAITLLSAGITDVCIEYKADDDGGTWRDNTYPGLTCDVPSRLYQY 64

QY 73 NIPEVWKDWTSWCRYPDQKELLSYVHHCDKIRGLKDVYFGAEVVDARYDLGTWTVKT 132
D 65 SFA---KNPNWTQMFSGGEIQDYLRGIAERYGLRHRIRFGATVVSARF--DDGRWVLR 119

QY 133 SAGHVATAKYLI L ATGLLH R KHTPALPGLADFN GKVIHSSAWHEDFDAEQRVAVIGAGA 192
D 120 DSGTESTVDFLISATGVLHHPRIPIPIAGLDLDFRGTVFHSARWDHTVPLLGRRIAVIGTGS 179

QY 193 TSIQIVQELAKKADQVTFMWRPSCYCLPM--RQRTMDRNEQTAWKAYYPTLFEASRKS 249
D 180 TGVQLVCGLAGVAGKVTMFQRTAQWVLPWPNPRYSKRLARVHRAFPCLGSLAYKAYSLSF 239

QY 250 IGEPVQAPSVGIFEVSPEQREAYFEELWERGAFNFLACQYREVVMVDKKANRLVYDFWAKK 309
D 240 ETFAVALSNPGL-----HRKLVGAVCRA 262

QY 310 TRSRIVNPAKRDLMAPLEPPYWFGTGRSPLESDDYVEMLDKPSVEIVNLEQSPIVAVTKTG 369
D 263 SLRRVRDPRLRALTPDYEPM--CKRLVMSGGFYRAIQRDDVELVT---AGIDHVEHRG 316

QY 370 VLLSDGSKRECDTIIVLATGFDSTFTGSLTHMGLKNKHGVDLKEVWKDGISTYMGVFESHGFP 429
D 317 IVTDDGVLHEVDVIVLATGFDSDH-AFFRPMQLTGRDGI RIDDVWQDGFPHAHQTVAIPGFP 375

QY 430 NAFVATAQAAPTIVLSNGP--TIETQVDLIADTIAKLEAEHATSVEATKSAQEAWSIMIA 487
D 376 NFFMMLGPHSP--VGNFPLTAVAESQAEHIVQWIKRWRHGEFDTMEPKSAATEAYNTVLR 433

QY 488 KMNEHTLFPL-TDSWW 502
D 434 AAMPNTVWTTGCDSWY 449

RESULT 3
Y4ID_RHISN
ID Y4ID_RHISN STANDARD; PRT; 662 AA.
AC P55487;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Probable monooxygenase Y4ID (EC 1.14.13.-).
GN Y4ID.
OS Rhizobium sp. (strain NGR234).
OC Plasmid sym pNGR234a.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
```



```
Db 261 -GAF-----YRDSMAQKASAFF-----RORMASIIKDDRLAKGFTPTFGGCRRI 305
QY 338 PLESDDYEMLDKPSVEIVNLEQSPIVAVTKTGVLLSDGSKRECDTIVLATGFD-SFTGSL 396
Db 306 TPGDPYMHAIQANVDV---HFTAVASCTEDGIVGADGIERLVDITVCASGFDNTYRPGF 362
QY 397 THMGLKNKHGVDLKEVWKDGISTYMGVFSHGFPNFAFFVATAQAPTVLSNGPTIETQVDL 456
Db 363 PIIG---RRGVDLRDKWKTNPAYLGLAVDMPN--YITFIGSPWPIONG----- 407
QY 457 IADTIKLEAEHATSVEATKSAQ-----EAWS---IMIAKMNEH 492
Db 408 --SVMAPLHVSVEYAIQFLKKMQNENIRAWAPROQITDRFNEH 448

RESULT 5
FMO2_MACMU STANDARD; PRT; 534 AA.
AC Q28505;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Dimethylaniline monooxygenase [N-oxide forming] 2 (EC 1.14.13.8)
DE (Pulmonary flavin-containing monooxygenase 2) (FMO 2) (Dimethylaniline
DE oxidase 2) (FMO 1B1).
GN FMO2.
OS Macaca mulatta (Rhesus macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
OC Cercopitheciinae; Macaca.
OX NCBI_TaxID=9544;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97214641; PubMed=9061021;
RA Yueh M.F., Krueger S.K., Williams D.E.;
RT "Pulmonary flavin-containing monooxygenase (FMO) in rhesus macaque:
RT expression of FMO2 protein, mRNA and analysis of the cDNA.";
RL Biochim. Biophys. Acta 1350:267-271(1997).
CC -!- FUNCTION: THIS PROTEIN IS INVOLVED IN THE OXIDATIVE METABOLISM OF
CC A VARIETY OF XENOBIOTICS SUCH AS DRUGS AND PESTICIDES.
CC -!- CATALYTIC ACTIVITY: N,N-dimethylaniline + NADPH + O(2) -> N,N-
CC dimethylaniline N-oxide + NADP(+) + H(2)O.
CC -!- COFACTOR: FAD AND MAGNESIUM ION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Microsomal.
CC -!- SIMILARITY: BELONGS TO THE FMO FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U59453; AAB02939.1; -.
DR InterPro; IPR000960; FMO.
DR Pfam; PF00743; FMO-like; 1.
DR PRINTS; PR00370; FMOXYGENASE.
KW Oxidoreductase; Monooxygenase; NADP; Flavoprotein; FAD; Microsome;
KW Transmembrane; Multigene family; Acetylation; Magnesium.
FT INIT_MET 0 0 BY SIMILARITY.
FT MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
FT NP_BIND 8 13 FAD (ADP PART) (POTENTIAL).
FT NP_BIND 190 195 NADP (POTENTIAL).
SQ SEQUENCE 534 AA; 60830 MW; B77D6F03B07CDB59 CRC64;

Query Match 7.6%; Score 220; DB 1; Length 534;
Best Local Similarity 21.8%; Pred. No. 1.3e-09;
Matches 103; Conservative 71; Mismatches 194; Indels 104; Gaps 15;

QY 19 IVGAGLSGISAVYKLRKLRLNAKIFEGAPDFGGVHHNRYPGARVDSETPFYQLNIPEVW 78
```

```
Db 6 VIGAGVSLISLKCVCDEGLEPTCFERTEDIGGVW---RFKEKVEDGRASIYQSVVTNTS 62
QY 79 KDWTSCHRYPDQKELLSYVHHCDKIRGLR-----KDYVGAENVVDARYDL---G 126
Db 63 KEMSCFSDFPMPEDFPNLFHNSKLLLEYERIPAKKFDLLKYIQFTVLSVRKCPDFSSSG 122
QY 127 TWTVKT--SAGHVATAKY--LILATGLLHRKHHTP--ALPGLADENGKVIHSSAWHEDFAE 181
Db 123 QWKVVTQSGNGKEQSAVFDVAVMVCVTHHFLPHILKSFPGIERFKQYFHSRQYKHPDGF 182
QY 182 GQVAVVIGACATSIQIVQELAKKADQVTMFMRPSSYCLPMRQRTMDRNEQTAWKAYPTL 241
Db 183 GKRLVIGMNGSGSDIAVELSKSAQVFISTRHGTWV----- 220
QY 242 FEASRKSRIGFPVQAPSVGIFEVSPQREAYFEELWERGAFNFLACQYREVMDKKANRL 301
Db 221 ---SRVSEGGYP-----WDSVFHTRFRSMLRNVLPRTVVVKWMIEQQMNQ- 261
QY 302 VYDFWAKKTRSRIVNPAKRDLMAPLEPPYWFCTKRSPLSDYEMLDKPSVEIVNLEQSP 361
Db 262 ---WPNHEN-----YGLEPQNKYIMKEPVLNDDVPSRLLCGAIKV---KST 301
QY 362 IVAVTKTGVLSDGSKRE-CDTIVLATGDFDSFTGSLTHMGLKNKHGVDLKEVWKDGISTY 420
Db 302 VKELTETSAIFEDGTVEENIDVVIATGY-----SFSFPFLED---SLVKVENNMVSLY 352
QY 421 MGVSFHGFPNFAFFVATAQAPTVLSNGPTIETQVDLIADTIKLEAEHATSV 472
Db 353 KYIFPAHLEKSTFACIGLIQLPLGSIFPT-----AEQARWVTV 391

RESULT 6
FMO2_CAVPO STANDARD; PRT; 534 AA.
AC P36366; Q05194;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Dimethylaniline monooxygenase [N-oxide forming] 2 (EC 1.14.13.8)
DE (Pulmonary flavin-containing monooxygenase 2) (FMO 2) (Dimethylaniline
DE oxidase 2) (FMO 1B1).
GN FMO2 OR FMO-2.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
OX NCBI_TaxID=10141;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=HARTLEY; TISSUE=Lung;
RX MEDLINE=93306345; PubMed=1306120;
RA Nikbakht K.N., Lawton M.P., Philpot R.M.;
RT "Guinea pig or rabbit lung flavin-containing monooxygenases with
RT distinct mobilities in SDS-PAGE are allelic variants that differ at
RT only two positions.";
RL Pharmacogenetics 2:207-216(1992).
CC -!- FUNCTION: THIS PROTEIN IS INVOLVED IN THE OXIDATIVE METABOLISM OF
CC A VARIETY OF XENOBIOTICS SUCH AS DRUGS AND PESTICIDES.
CC -!- CATALYTIC ACTIVITY: N,N-dimethylaniline + NADPH + O(2) -> N,N-
CC dimethylaniline N-oxide + NADP(+) + H(2)O.
CC -!- COFACTOR: FAD AND MAGNESIUM ION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Microsomal.
CC -!- TISSUE SPECIFICITY: LUNG.
CC -!- POLYMORPHISM: THERE ARE TWO ALLELIC FORMS (A AND B).
CC -!- SIMILARITY: BELONGS TO THE FMO FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```



```
QY 79 KDWTSCTRYDQKELLSYVHCHDKIRGLR-----KDYFGEAEVVDARYARDL---G 126
|: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 63 KEMSCFSDFPMPEDFPNLFHNSKLLLEYFRIFAKKFDLLKYIQFTTVLSVRKCPDFSSG 122

QY 127 TWTVKT-SAGHVATKY--LILATGLLRKHTP--ALPGLADFNKGKVIHSSAWHEDFDAE 181
|: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 123 QWKVVTSQNGKEQSAVFDVAVMVCSGHHILPHILKSPFGMERFKGQYFHSRQYKHPDGFE 182

QY 182 QORVAVIGAGATSIQIVQELAKKADQVTMFMRRPSYCLPMRQRTMDRNEQTAWKAYYPTL 241
|: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 183 GKRLVIGMNGSGSDIAVELSKNAQVFISTRHGTWVM----- 220

QY 242 FEASRKSRIQFPVQAPSVGIFEVSPEQREAYFEELWERGAFNFLACQYREVMVDKKANRL 301
|: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 221 ---SRISEDGYP-----WDSVFHTFRSMLRNVLPRPTAVKWMIEQQMNR- 261

QY 302 VYDFWAKKTRSRIVNPAKRDLMAPLEPPYWGFKRSPLESDDYEMLDKPSVEIVNLEQSP 361
|: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 262 ----WFNHEN-----YGLEPQNKYIMKEPVLNDDVPSRLLCGAIV----KST 301

QY 362 IVAVTKTGVLSDGSKRE-CDTIVLATGDFSFT 393
|: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 302 VKELTETSATFEDGTVEENIDVIFATGY-SFS 333

RESULT 11
FMO2_RABIT STANDARD; PRT; 534 AA.
AC P17635;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Dimethylaniline monooxygenase [N-oxide forming] 2 (EC 1.14.13.8)
DE (Pulmonary flavin-containing monooxygenase 2) (FMO 2) (Dimethylaniline
DE oxidase 2) (FMO 1B1).
GN FMO2.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=NEW ZEALAND WHITE;
RX MEDLINE=90202836; PubMed=2318837;
RA Lawton M.P., Gasser R., Tynes R.E., Hodgson E., Philpot R.M.;
RT "The flavin-containing monooxygenase enzymes expressed in rabbit
RT liver and lung are products of related but distinctly different
RT genes.";
RL J. Biol. Chem. 265:5855-5861(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NEW ZEALAND WHITE; TISSUE=Lung;
RX MEDLINE=93306345; PubMed=1306120;
RA Nikbakht K.N., Lawton M.P., Philpot R.M.;
RT "Guinea pig or rabbit lung flavin-containing monooxygenases with
RT distinct mobilities in SDS-PAGE are allelic variants that differ at
RT only two positions.";
RL Pharmacogenetics 2:207-216(1992).
RN [3]
RP PARTIAL SEQUENCE, AND ACETYLATION.
RC TISSUE=Lung;
RX MEDLINE=92002038; PubMed=1911780;
RA Guan S., Falick A.M., Williams D.E., Cashman J.R.;
RT "Evidence for complex formation between rabbit lung flavin-containing
RT monooxygenase and calreticulin.";
RL Biochemistry 30:9892-9900(1991).
CC -!- FUNCTION: THIS PROTEIN IS INVOLVED IN THE OXIDATIVE METABOLISM OF
CC A VARIETY OF XENOBIOTICS SUCH AS DRUGS AND PESTICIDES.
CC -!- CATALYTIC ACTIVITY: N,N-dimethylaniline + NADPH + O(2) - N,N-
CC dimethylaniline N-oxide + NADP(+) + H(2)O.
CC -!- COFACTOR: FAD AND MAGNESIUM ION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Microsomal.
```

```
CC -!- TISSUE SPECIFICITY: LUNG.
CC -!- POLYMORPHISM: THERE ARE TWO ALLELIC FORMS.
CC -!- SIMILARITY: BELONGS TO THE FMO FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M32029; AAA31442.1; -.
CC PIR; B35182; B35182.
CC InterPro; IPR000960; FMO.
CC Pfam; PF00743; FMO-like; 1.
CC PRINTS; PR00370; FMOXYGENASE.
CC Oxidoreductase; Monooxygenase; NADP; Flavoprotein; FAD; Microsome;
CC Transmembrane; Multigene family; Acetylation; Magnesium; Polymorphism.
CC INIT_MET 0 0
CC MOD_RES 1 1 ACETYLATION.
CC NP_BIND 8 13 FAD (ADP PART) (POTENTIAL).
CC NP_BIND 190 195 NADP (POTENTIAL).
CC VARIANT 119 119 A -> S (IN FMO-2 AND FMO-4).
CC VARIANT 135 135 Q -> E (IN FMO-2 AND FMO-4).
CC SEQUENCE 534 AA; 61012 MW; E7066BDA41C4135 CRC64;

Query Match 7.2%; Score 207.5; DB 1; Length 534;
Best Local Similarity 20.2%; Pred. No. 1.2e-08;
Matches 103; Conservative 81; Mismatches 193; Indels 133; Gaps 17;

QY 19 IVGAGLSGISAVYKRLRLNAKIFEGAPDFGGVWHNRYPGARVDSETPFYQLNIPEVW 78
|: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 6 VIGAGVSLISLKCVCDEGLEPTCFERTEDIGGLW---RFKENVEDGRASIQSVITNTS 62

QY 79 KDWTSCTRYDQKELLSYVHCHDKIRGLR-----KDYFGEAEVVDARYARDL---G 126
|: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 63 KEMSCFSDFPMPEDFPNLFHNSKLLLEYFRIFAKKFDLLKYIQFTTVLSVRKCPDFASSG 122

QY 127 TWTVKT---SAGHVATKYILATG--LLHRKHTPALPGLADFNKGKVIHSSAWHEDFDAE 181
|: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 123 QWEVVTQSNKQSAVDVAVMVCSGHHILPNILKSPFGIEKFKGOYFHSRQYKHPAGLE 182

QY 182 QORVAVIGAGATSIQIVQELAKKADQVTMFMRRPSYCLPMRQRTMDRNEQTAWKAYYPTL 241
|: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 183 GKRLVIGIGNSASDIAVELSKKAAQVYISTRKGSWM----- 220

QY 242 FEASRKSRIQFPVQAPSVGIFEVSPEQREAYFEELWERGAFNFLACQYREVMVDKKANRL 301
|: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 221 ---SRISEDGYP-----WDMVFHTFRSMLRNVL-----PRM 249

QY 302 VYDFWAKKTRSRIVN-----PAKRDLMAPLEPPYWGFKRSPLESDDYEMLDKPS--- 351
|: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 250 IVKWMIEQQMNRWENHENYGLAPENKYL-----KEPVLND-----DLPSRIL 292

QY 352 ---VEIVNLEQSPIVAVTKTGVLSDGS-KRECDTIVLATGDFSFTGSLTHMGLKNKHGV 407
|: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 293 YGTIKV---KRRVKELTESAAIFEDGTVEEDIVVFATGYTFAFPFLEE----- 339

QY 408 DLKEVWKDGISTYMGVFSHGFPNAPFVATAQAPTIVLSNGPTTIETQVDLIADTIKLEAE 467
|: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 340 SLVKIEDNMVSLYKYMFPFPQLEKSTFACGLIQLPLGSIFFTV-----ELQAR 386

QY 468 HATSV---EATKSAQEAWSIMIAKMNEHTL 494
|: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 387 WATRVFKGLCSLPSKETMMADIKRNNRI 416

RESULT 12
Y943_MYCTU
ID Y943_MYCTU STANDARD; PRT; 346 AA.
AC P71566;
```



```
Db 250 -----ISDWWYVKQMAKFKHENYSLM-PLN-----GTLRKEPVFND-----DLP 288
QY 351 S----VEIVNLEQSPIVAVTKTVLLSDGSKREC-DTIVLATGFDSDFTGSLTHMGLKNKHG 406
Db 289 ARILCGTVSIKPN-VKEFTETSAIFEDGTVFEAIDSVIFATGYGYAYPFLDDSIKSEN- 346
QY 407 VDLKEVWKDGISTYMGVFSHGFPNFAFVATAQAPTVLSNGPTTIETQVDLIADTIAKLEA 466
Db 347 -----NKVTLFKGIF-----PPQLEKPTMAVIGLVQSLGA 376
QY 467 EHATSVKATKSAQEAWSIMIAKMNEHTLFLPLTD 499
Db 377 AIPPT-----DLQARWAAQVIK-GTCTLPPVKD 403

RESULT 15
FMO3_MOUSE
ID FMO3_MOUSE STANDARD; PRT; 534 AA.
AC P97501;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Dimethylaniline monooxygenase [N-oxide forming] 3 (EC 1.14.13.8)
DE (Hepatic flavin-containing monooxygenase 3) (FMO 3) (Dimethylaniline
DE oxidase 3).
DE
DE
GN Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CD-1; TISSUE=Liver;
RX MEDLINE=98008021; PubMed=9344459;
RA Falls J.G., Cherrington N.J., Clements K.M., Philpot R.M., Levi P.E.,
RA Rose R.L., Hodgson E.;
RT "Molecular cloning, sequencing, and expression in Escherichia coli of
RT mouse flavin-containing monooxygenase 3 (FMO3): comparison with the
RT human isoform.";
RL Arch. Biochem. Biophys. 347:9-18(1997).
CC -|- FUNCTION: THIS PROTEIN IS INVOLVED IN THE OXIDATIVE METABOLISM OF
CC A VARIETY OF XENOBIOTICS SUCH AS DRUGS AND PESTICIDES. FORM II
CC N-OXYGENATES PRIMARY ALIPHATIC ALKYLAMINES AS WELL AS SECONDARY
CC AND TERTIARY AMINES (BY SIMILARITY).
CC -|- CATALYTIC ACTIVITY: N,N-dimethylaniline + NADPH + O(2) = N,N-
CC dimethylaniline N-oxide + NADP(+) + H(2)O.
CC -|- COFACTOR: FAD (BY SIMILARITY).
CC -|- SUBCELLULAR LOCATION: Microsomal.
CC -|- SIMILARITY: BELONGS TO THE FMO FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U87147; AAB47541.1; -
DR MGD; MGI:1100496; Fmo3.
DR InterPro; IPR000960; FMO.
DR Pfam; PF00743; FMO-like; 1.
DR PRINTS; PR00370; FMOXYGENASE.
KW Oxidoreductase; Monooxygenase; NADP; Flavoprotein; FAD; Microsome;
FT NP_BIND 9 14 FAD (ADP PART) (POTENTIAL).
FT NP_BIND 191 196 NADP (POTENTIAL).
SQ SEQUENCE 534 AA; 60516 MW; F72F7993C01AF9C9 CRC64;
```

Query Match 6.8%; Score 198; DB 1; Length 534;
Best Local Similarity 21.2%; Pred. No. 6.9e-08;

```
Matches 103; Conservative 69; Mismatches 183; Indels 130; Gaps 19;
QY 19 IVGAGLSGISAVYKLRKLRNAKIFEGAPDFGVVWNNRYPGARVDSETPFYQLNIPVW 78
Db 7 IIGAGVSGLAIRSCLEEGLEPTCFERSDDVGGLW---KFSDHIEGRASIYQSVFTNSS 63
QY 79 KDWTWSCRYPDQKELLSYVHCD-----KIRGLRKDYVFGAEVVDARYARDL--G 126
Db 64 KEMMCFDPDFYPDDFPNFMHHSKIQEYITSAKEKNLLKYIOFETPVTINKCPNFSTTG 123
QY 127 TWTVKTSAGHVATAKY--LILATGLLHRKHTP--ALPGLADFNKGVHSSAWHEDFAE 181
Db 124 KWEVTEKHGKKETAVIDATMICSGHIIFFHPVKDSFPGLNRFKGCFCFHSRDYKEPGIWK 183
QY 182 GQRVAVIGACATSIQIVQELAKKADQVTFMRRRPSYCLPMRQRTMDRNEQTAWKAYYP-- 239
Db 184 GKRVLVIGLNSGCDIAAELSHVAQKVTISSRSGSWM-----SRWDDGYPWD 232
QY 240 ----TLFEASRKSRIGFPVQAPSVGIFEVSPEQREAYFEELWERGAENFLACQYREVMVD 295
Db 233 MVLTRFQTLKNNL-----PTA----- 250
QY 296 KKANRLVYDFW-AKKTISRIVN-----PAKRDLMAPLEPPYWFGRKSPLESYYEMLD 348
Db 251 -----ISDWWYTRQMNARFKHENYGLVPLNRTL-----RKEPVFND--ELPA 290
QY 349 KPSVEIVNLEQSPIVAVTKTVLLSDGSKREC-DTIVLATGFDSDFTGSLTHMGLKNKHG 407
Db 291 RILCGMVTIKPN-VKEFTETSAVEDGTMFEAIDCVIFATGYGYAYPFLDDSIKSRN-- 347
QY 408 DLKEVWKDGISTYMGVFSHGFPNFAFVATAQAAPTIVLSNGPTTIETQVDLIADTIAKLEA 467
Db 348 -----NEVTLYKGVEP---PQLEKPTMAVIGLVQSLGATI-----PITDLQAR 387
QY 468 HATSV 472
Db 388 WAAQV 392
```

Search completed: April 27, 2002, 07:59:37
Job time: 743 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2002, 05:56:13 ; Search time 68.15 Seconds
(without alignments)
768.433 Million cell updates/sec

Title: US-09-882-694A-3
Perfect score: 2902
Sequence: 1 MSATNSNRGDCSVACDAIIV.....QICQEKVANWDGFDVLHAPC 545

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues 283138
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	895.5	30.9	549	2 JC7158	steroid monooxygen
2	838	28.9	556	2 C87472	steroid monooxygen
3	802	27.6	542	2 A28550	cyclohexanone mono
4	786	27.1	543	2 E87474	steroid monooxygen
5	550.5	19.0	491	2 F83383	probable flavin-bi
6	509.5	17.6	492	2 G70899	probable monooxyge
7	500.5	17.2	495	2 A70782	probable monooxyge
8	497.5	17.1	524	2 D70861	probable monooxyge
9	466.5	16.1	527	2 A83453	probable flavin-co
10	362.5	12.5	489	2 C70655	probable monooxyge
11	359	12.4	499	2 E87416	monooxygenase, fla
12	346.5	11.9	494	2 A86917	probable monooxyge
13	342.5	11.8	499	2 B83493	conserved hypothet
14	339	11.7	833	2 G75621	arylesterase/monox
15	337	11.6	498	2 G87567	monooxygenase, fla
16	326	11.2	486	2 D70932	probable monooxyge
17	297	10.2	495	2 G70852	hypothetical prote
18	232	8.0	431	2 H84742	probable flavin-co
19	219	7.5	458	2 C87345	monooxygenase, fla
20	216.5	7.5	414	2 T08587	hypothetical prote
21	216.5	7.5	533	2 S71617	dimethylaniline mo
22	216	7.4	530	2 B86326	T29M8.12 protein -
23	213.5	7.4	533	2 A46677	dimethylaniline mo
24	212	7.3	426	2 T04527	hypothetical prote
25	212	7.3	528	2 T22583	hypothetical prote
26	210	7.2	421	2 T00955	hypothetical prote
27	210	7.2	514	2 A45912	flavin-containing
28	210	7.2	514	2 A54250	microsomal flavin
29	207.5	7.2	535	1 B35182	dimethylaniline mo

30	206.5	7.1	346	2 E70715	hypothetical prote
31	206.5	7.1	548	2 T23436	hypothetical prote
32	202.5	7.0	437	2 G86178	hypothetical prote
33	202.5	7.0	533	2 S71618	dimethylaniline mo
34	200.5	6.9	531	2 B54096	flavin-containing
35	200	6.9	469	2 AG3139	flavin-containing
36	200	6.9	605	2 E98148	hypothetical prote
37	199	6.9	458	2 T37052	probable flavin-co
38	197.5	6.8	533	2 S51131	flavin-containing
39	197	6.8	406	2 T36632	probable oxidoredu
40	194.5	6.7	529	2 T23431	hypothetical prote
41	192.5	6.6	558	2 S29125	dimethylaniline mo
42	189.5	6.5	555	2 A54096	flavin-containing
43	188.5	6.5	600	2 G95854	probable monooxyge
44	187	6.4	415	2 T07706	hypothetical prote
45	181	6.2	533	1 S18380	dimethylaniline mo

ALIGNMENTS

RESULT 1

JC7158
steroid monooxygenase (EC 1.14.99.-) - Rhodococcus rhodochrous
C;Species: Rhodococcus rhodochrous
C;Date: 04-Mar-2000 #sequence_revision 04-Mar-2000 #text_change 21-Jul-2000
C;Accession: JC7158; PC7046
R;Morii, S.; Sawamoto, S.; Yamauchi, Y.; Miyamoto, M.; Itagaki, E.
J. Biochem. 126, 624-631, 1999
A;Title: Steroid monooxygenase of Rhodococcus rhodochrous: Sequencing of the genomic
A;Reference number: JC7158; MUID:99398443
A;Accession: JC7158
A;Molecule type: DNA
A;Residues: 1-549 <MOR>
A;Cross-references: DDBJ:AB010439; NID:g2804297; PIDN:BAA24454.1; PID:g2804298
A;Experimental source: strain IFO 3338
A;Accession: PC7046
A;Molecule type: protein
A;Residues: 1-15;70-87;95-114;172-181;194-211;312-319;335-352;413-422;473-492;518-53
C;Comment: This enzyme is a Baeyer-Villigerase catalyzing the insertion of an oxyge
ble protein having catalytic properties.
C;Genetics:
A;Gene: smo
A;Start codon: TTG
C;Keywords: FAD; monooxygenase; nucleotide binding; oxidoreductase

Query Match 30.9%; Score 895.5; DB 2; Length 549;
Best Local Similarity 36.3%; Pred. No. 3.6e-62;
Matches 192; Conservative 103; Mismatches 227; Indels 7; Gaps 6;

QY	16	DAIIVGAGLSGISAVYKLRKLRNLNAKIFEGAPDFGGVWHWNRYPGARVDSPTFFYQLNI- 74
		: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db	23	DVVVVGACIAGLYAIHRFRSQGLTVRAFEAASGVGVWYWNRYPGARCDVESIDYSYFS 82
		: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY	75	PEVWKDWTWSCRYPDQKELLSYVHHCDKIRGLRKDVYFGEAEVVDARYARDLGTWTVKTS 134
		: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db	83	PELEQEWNWSEKYATQPEILAYLEHVADRFDLRRDIRFTRVTSAVLDEEGLRWTVRTDR 142
		: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY	135	GHVATAKYLIILATGLLRKHKHTPALPLGADPNKGVHSSAW-HEDFDAEQQRVAIVIGAGAT 193
		: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db	143	GDEVSARPLVVAAGPLSNANTPAFDGLDRFTGDIVHTARWPHDGVDFTKRGVGVIGTGSS 202
		: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY	194	SIQIVQELAKKADQVTFMFMRRRPSYCLPMRQTFMDRNEQTAKWAYPTLFEASRKSRIQFP 253
		: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db	203	GIQSIPITAEQAEQLFVFORANSYISIPAGNVPLDDATRAEQKANYAERRRLSRSGGGSP 262
		: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY	254	VQAPSVGIFEVSPQREAYFEELWERGAFNFLACQYREVMDKKANRLVYDFWAKKTRSR 313
		: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db	263	HRPHPKSALEVSEERRAVYEERWKLGGVLF-SKAPPDQLTDPAAANDTARAFWEKIRAV 321
		: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY	314	IVNPAKRDLMAPLEPPYWFCTKRSPLSDYYEMLDKPSVEIVNLEQSPIVAVTKTGVLSS 373
		: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Db 322 VDDPAVELLTPKD--HAIGAKRIVTDSGYETYNRDNVELDLRSTPIVGMDETGI-VT 378

QY 374 DGSKRECDTIVLATGDFSFTGSLTHMGLKNKHGVDLKEVWKDGISTYMGVFSHGFPNAFF 433

Db 379 TGAHYDLDMIVLATGFDAMTGSCLKLEIVGRGGRTLKETWAAGPRTYVLGLGIDGFPNFFN 438

QY 434 VATAQAPTIVLSNGPTIETQVDLIADTIAKLEAEHATSVEATKSAQEAWSIMIKNHEHT 493

Db 439 LTGPGSPSVLANMVLHSELHVDWADAIAYLDAARGAAGIEGTPEAVADWVEECRNRAEAS 498

QY 494 LFPLTDSWWTGGNIPGKATRALTFIGGIALYEQICQEKV-ANWDDGFDVL 541

Db 499 LLNSANSWYLGANIPGRPRVFMPLGGFGVYRELIITEVAESGYKGFALL 547

RESULT 2

C87472

steroid monooxygenase [imported] - Caulobacter crescentus

C;Species: Caulobacter crescentus

C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001

C;Accession: C87472

R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J. B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M. Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A;Title: Complete Genome Sequence of Caulobacter crescentus.

A;Reference number: A87249; MUID:21173698; PMID:11259647

A;Accession: C87472

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-556 <STO>

A;Cross-references: GB:AE005673; NID:g13423231; PIDN:AAK23775.1; GSPDB:GN00148

C;Genetics:

A;Gene: CCI799

Query Match 28.9%; Score 838; DB 2; Length 556;

Best Local Similarity 33.4%; Pred. No. 1.2e-57;

Matches 179; Conservative 104; Mismatches 235; Indels 18; Gaps 6;

QY 16 DAIIVGAGLSGISAVYKLR-LRLNNAKIFEGAPDFGGVWHNNRYPGARVDSETPFYQLNI- 74

Db 17 DALVIGAGFGGMYAVHKLSCGMSVQGFAGGDDVGGVWYNNRYPGARCDLMSLDYCYSES 76

QY 75 PEVWKDWTWSCRYPDQKELLSYVHHCDKIRGLRKDVFGAEVVDARYARDLGTWTVKTS 134

Db 77 PDIEQEWWTSEQFAAQSEILAYANFVADRLDLRRHFLFETWVTGAAYDEAANVWRVTVD 136

QY 135 GHVATAKYLILATGLLHRKHTPALPLGADFNKGVIHSSAW-HEDFDAEQRVAVIGACAT 193

Db 137 GRVFEATYCVMASGPLSTPKGVPPFDGADDFKGEILLAAKWPHTPVSFEGKRVGLVGTST 196

QY 194 SIQIVQELAKKADQVTMFMRPPSYCLPMRQRTMDRNEQTAWKAYPTLFEASRKSRIGFP 253

Db 197 GIQIVPEVAKTAGALTVPQRTPSFTLPMRNVTLPEPDYVAELKRNAGIRQVARNSPGLGV 256

QY 254 VQAPSVGIFEVSPEQRAYFEELWERGAFNFLACQYREVWVDKKNRLLVYDFWAKKTRSR 313

Db 257 RPQSTRPFFSVTPQORALMEDSWARGGLSFLG-TFSDLLVNPEANEQVAEEFVRSKISEV 315

QY 314 IVNPAKRDLMAPLEPPYFPGTKRSPLESDDYYEMLDKPSVEIVNLEQSPIVAVTKGVLLS 373

Db 316 VTDPVTAELKPRGYPF--ARRPCLDTHYYETENLPHVVKLHDCLESPIVKITPAG-LLT 372

QY 374 DGSKRECDTIVLATGDFSFTGSLTHMGLKNKHGVDLKEVWKDGISTYMGVFSHGFPNAFF 433

Db 373 ETGEVELDVLIATGYDGLTGALLNFDVVGRGGLQLRDKWRDGAKSHLGLMIASFPNLF 432

QY 434 VATAQAPTIVLSNGPTIETQVDLIADTIAKLEAEHATSVEATKSAQEAWSIMIKNHEHT 493

Db 433 VCGPNGPAALANITLDQQNIDWIADAITHMRQEGLATIEPTEGAEQGWMDMYVGMAELT 492

QY 494 LFPLTDSWWTGGNIPGKATRALTFIGGIALYEQICQ-----EKVANWDG 537

Db 493 LVSKANTWYVGGNISGKPRGLSMYTGGFQRYSEACRLAAEKGYEDFVEKAATKQG 548

RESULT 3

A28550

cyclohexanone monooxygenase (EC 1.14.13.22) - Acinetobacter sp.

C;Species: Acinetobacter sp.

C;Date: 08-Mar-1989 #sequence_revision 08-Mar-1989 #text_change 08-Oct-1999

C;Accession: A28550; B28550

R;Chen, Y.C.J.; Peoples, O.P.; Walsh, C.T.

J. Bacteriol. 170, 781-789, 1988

A;Title: Acinetobacter cyclohexanone monooxygenase: gene cloning and sequence deteri

A;Reference number: A28550; MUID:88115180

A;Contents: NCIB 9871

A;Accession: A28550

A;Molecule type: DNA

A;Residues: 1-542 <CH1>

A;Cross-references: GB:M19029; NID:g141767; PIDN:AAA21892.1; PID:g141768

A;Accession: B28550

A;Molecule type: protein

A;Residues: 2-11 <CH2>

C;Comment: The initiator Met is not shown.

C;Keywords: oxldoreductase

Query Match 27.6%; Score 802; DB 2; Length 542;

Best Local Similarity 33.4%; Pred. No. 7.6e-55;

Matches 179; Conservative 96; Mismatches 237; Indels 24; Gaps 8;

QY 16 DAIIVGAGLSGISAVYKLR-KLRLNAKIFEGAPDFGGVWHNNRYPGARVDSETPFY---- 70

Db 7 DAIVIGGGFGGLYAVKKLRDELELKVQAFDKATDVAGTWYNNRYPGALTDTETHLYCYSW 66

QY 71 ---QLNIPEVWKDWTWSCRYPDQKELLSYVHHCDKIRGLRKDVYFGAEVVDARYARDLGT 127

Db 67 DKELLOSLEIKK-----KYVQGPDRVRYLQQAEEKHDLKKSQFNTAVQSAHYNEADAL 120

QY 128 WTVKTSAGHVATAKYLILATGLLHRKHTPALPLGADFNKGVIHSSAWHEDFDAEQRVAV 187

Db 121 WEVTTEYGDKYTARPLITLGLLSAPNLPNIKGINQPKELHHTSRWPDVSVFEGKRVGV 180

QY 188 IGAGATSIQIVQELAKKADQVTMFMRPPSYCLPMRQRTMDRNEQTAWKAYPTLFEASRK 247

Db 181 IGTGSTGVQVITAVAPLAKHLTVFQSAQYSVPIGNDPLSEEDVKKIKDNYDKSLGWCNM 240

QY 248 SRIGFPVQAPSVGIFEVSPEQRAYFEELWER-GAFNFLACQYREVWVDKKNRLLVYDFW 306

Db 241 SALAFALNESTVPAMSVSAEERKAVFEKAWQTGGGFRFMFETFGDIATNMEANIEAQNFI 300

QY 307 AKTRSRIVNPAKRDLMAPLEPPYFPGTKRSPLESDDYYEMLDKPSVEIVNLEQSPIVAVT 366

Db 301 KGKIAEIVKDPATAQKLMPOD---LYAKRPLCDSGYNTFNDRNVRLEDVKANPIVEIT 356

QY 367 KTGVLSDGSKRECDTIVLATGDFSFTGSLTHMGLKNKHGVDLKEVWKDGISTYMGVFSH 426

Db 357 ENGVKLENGDFVELDMLICATGFDVADGNYVVRMDIQKNGLAMKDYWKEGPSSYMGVTVN 416

QY 427 GFPNAFFVATAQAPTIVLSNGPTIETQVDLIADTIAKLEAEHATSVEATKSAQEAWSIMI 486

Db 417 NYPNMFVVLGPNGP--FTNLPPSIESQVEWISDTIQYTVNNVNESIEATKEAEEQWTQTC 474

QY 487 AKMNEHTLPLTDSWWTGGNIPGKATRALTFIGGIALYEQICQEKVAN--WDGFDV 540

Db 475 ANIAEMTLFPKAQSWIFGANIPGKKNTVYFYLGLKEY-RTCASCNCKNHAYEGFDI 529

RESULT 4

E87474

steroid monooxygenase [imported] - Caulobacter crescentus

C;Species: Caulobacter crescentus

C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001

C;Accession: E87474

Db 251 --LGIY-----WAEALAY-----GMTKRPNTLKIIEAYAKYNIRRSV 286

QY 316 NPAKRDLMAPLEPPYFPGTKRSPLESDYYEMLDKPSVEIVNLEQSPIVAVTKTGVLSDG 375

Db 287 K--DRELRRKLTPRYIGCKRIILNSSTYYPAVADPKTELIT---DRIDRITHDGIVTADG 341

QY 376 SK----RECDTVLATGDFSFTGSLTHMGLKNKHGVDLKEVW-KDGISTYMGVFSHGFPN 430

Db 342 TGREVFREADVIVATGF-HVTDSTYYVQIKGRHGEDLVDRWNREGIGAHRGITVANMPN 400

QY 431 AFFVATAQAFTVLSNGPTIIETQVDLIADTIAKLEAEHATSVEATKSAQEAWSIMIAKMN 490

Db 401 LFFLLGPNTGLGHNSVVFMIESQIHYYVADAIKCDRMGVQALAPTREAQD-----RFN 453

QY 491 EHTLPLTDSWWTGG 505

Db 454 QELQRRLAGSVVNSG 468

RESULT 9

A83453

probable flavin-containing monooxygenase PA1538 [imported] - Pseudomonas aeruginosa (str

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: A83453

R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,

.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho

A:Reference number: A82950; MUID:20437337

A:Accession: A83453

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-527 <STO>

A:Cross-references: GB:AE004582; GB:AE004091; NID:g9947492; PIDN:AAG04927.1; GSPDB:GN001

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PA1538

QY 423 VFSHGFPNAPFFVATAQAFTVLSNGPTIIETQVDLIADTIAKLEAEHATSVEATKSAQEAW 482

Db 392 TTTAGFPNLFPLMGPNNTGLGHNSVMVIESQIAYVLDALKMKRRELLSLEVKAPVQE-- 449

QY 483 SIMIAKMNEHTLFPPLTDSWWTGG 505

Db 450 -----RYNEYIQRKLDRSVWSVG 467

RESULT 10

C70655

probable monooxygenase - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: C70655

R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon

; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete gen

A:Reference number: A70500; MUID:98295987

A:Accession: C70655

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-489 <COL>

A:Cross-references: GB:Z83864; GB:AL123456; NID:g3261687; PIDN:CAB06212.1; PID:e301

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: Rv3854c

Query Match 12.5%; Score 362.5; DB 2; Length 489;

Best Local Similarity 26.3%; Pred. No. 2e-20;

Matches 124; Conservative 66; Mismatches 187; Indels 95; Gaps 18;

QY 16 DAIIVGAGLSGISAVYKLRKLRLNAK---IFEGAPDFGGVHHWNRYPCARVDSETPFYQL 72

Db 6 DVVIVGAGISGVSAAWHLQD-RCPTKSYAILEKRESMGGTWDLFRYPGIRSDSD--MYTL 62

QY 73 NIPEVWKDWTWSCRYPDQKELLSYVHHCDKIRGLRKDVYFGAEVVDARYARDLGTWTVK- 131

Db 63 GFR--FRPWTGRQAIADGKPILEYVKSTAAMYGIDRHIRFHHKVISADWSTAENRWTVHI 120

QY 132 TSAGHVA--TAKYLILATGLLH--RKHTPALPGLADFNKGKVIHSSAWHEDFDAEQRVAV 187

Db 121 QSHGTLALTCEFLFCSGYNYNDEGYSRPFAGSEDFVGPIIHQHPWPELDLDYDAKNIV 180

QY 188 IGAGATSIQIVQELAKK-ADQVTMFMRPPSYCLPMRQRTMDRNEQTAWKAYYPTLFEASR 246

Db 181 IGSGATAVTLVPALADSGAKHVTMLQRSPTYIVSQPDRDGAIEKLRW----- 228

QY 247 KSRIGPPVQAPSVGIFEVSPEQREAYFEELWE--RGAFNELACQYREVMDKKANRLVY 303

Db 229 -----LPETM-AYTAVRWKNVLRQAAVYSACQK----- 255

QY 304 DFWAKKTRSRIVNPAKRDLMAPLEPPYFPGTKRSPLE-----SDYYEMLDKPSVEIV 355

Db 256 --WPRMRKMFSLIQORQLPEGYDVRKHFGPHYNPDQRLCLVPNGDLFRAIRHGKVEV 313

QY 356 NLEQSPIVAVTKTGVLSDGSKRECDTIVLATGFD--SFTGSLTHMGLKNKHGVDLKEVW 413

Db 314 T---DTIERFTATGIRLSNGRELPAIIITATGLNLQFGGATA-----TIDGQQVD 362

QY 414 KDGISTYMGVFSHGFPN-AFFVATAQAFTVLSNGPTIIETQVDLIADTIAKL 464

Db 363 ITTMMAYKGMMLSGIPNMAVTVGYTNASWTL-----KADLVSEFVCRL 405

RESULT 11

E87416

monooxygenase, flavin-binding family [imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus

C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C;Accession: E87416
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of *Caulobacter crescentus*.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: E87416
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-499 <STO>
A;Cross-references: GB:AE005673; NID:gl34222697; PIDN:AAK23329.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC1348

Query Match 12.4%; Score 359; DB 2; Length 499;
Best Local Similarity 24.6%; Pred. No. 3.8e-20;
Matches 123; Conservative 72; Mismatches 197; Indels 108; Gaps 18;

QY 7 SRGDCSVACDAIIIVGAGLSGISAIVYKLRKRLNAK--IFEGAPDFGGVWHNRYPGARVD 64
DB 11 AHGGATOHFDVIIVGAGISGIGAYHLTTQRPGRTRYCVLEALDSFGGTWLTHTYPGIRSD 70
QY 65 SETPFYQNLNIPEVWKDWTWSCR-----PDQKELLSYVHHCKIRGLRKDVYFGAEV 116
DB 71 SDL-----YTFGYRKPWPVGPPIATAAEILSYMGEVIDENGLAPHIRYRRI 117
QY 117 VDARYARDLGTWTVKTSAG---HVATAKYLILATGLLHRK--HTPALPGLADFNKGKVIHS 171
DB 118 SSARWSSKDKLTWLTQVTGPDGVETTYTNFLWMCQGYRHSVGYTPWPGMADFGCSIVHP 177
QY 172 SAWHEDFDAEGQORVAVIGAGATSIQIVQELAKKADQVTFMRRRPSYCLPMRQRTMDRNEQ 231
DB 178 QTWPADLDLKGKKVVVIGSGATAATLVPNIAGDCEHVTLQRSPTYFVP-----GRNEN 231
QY 232 TAWKAYPTLFEASRKSRIGFPVQAPSVGIFEVSPEQREAYFEELWERGAFNPLACQYRE 291
DB 232 -----ELANTLRQLOVDESWIHEIVR-----RK 254
QY 292 VMVDKK--ANRLVYDFWAKKTRSRIVNPAKR-----DLMAPLEPPY--WFGTKRSPLES 341
DB 255 VLFDOHEFTRRAIEE--SDTVKAEELLEGVKMFLEDGDFDVAKHFTPRYRPWRORIAFVPDG 312
QY 342 DYYEMLDKPSVEIVNLEQSPIVAVTKTGVLSDGSKRECDTIVLATGFD-SFTGSLTHMG 400
DB 313 DLFQGVASGKASVVTDE---IECFRTGLLKSGETLDADVIITATGFDLSVLGDIA--- 366
QY 401 LKNKHGVDLKEVWKDGISTYMGVFSHGFPNFAFFV-ATAOAPTVLNSNGPTIETQVDLIAD 459
DB 367 ----FEIDGQPLDFAKTVTYRGMFTGVPNLVWVFGYFRASWTL-----RADLIGD 413
QY 460 TIAKLEAEHATSVEATKSAQ 479
DB 414 FVCRLLAH-----MEKTGAKQ 429

RESULT 12
A86917
probable monooxygenase [imported] - *Mycobacterium leprae*
C;Species: *Mycobacterium leprae*
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C;Accession: A86917
R;Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Ho
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,
eam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A;Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; S
A;Title: Massive gene decay in the leprosy bacillus.
A;Reference number: A86909; MUID:21128732; PMID:11234002
A;Accession: A86917
A;Status: preliminary

A;Molecule type: DNA
A;Residues: 1-494 <STO>
A;Cross-references: GB:AL450380; NID:gl3092457; PIDN:CAC29573.1; GSPDB:GN00147
C;Genetics:
A;Gene: ML0065
Query Match 11.9%; Score 346.5; DB 2; Length 494;
Best Local Similarity 25.6%; Pred. No. 3.6e-19;
Matches 121; Conservative 70; Mismatches 185; Indels 97; Gaps 19;
QY 16 DAIIVGAGLSGISAVYKLRKRLNAK--IFEGAPDFGGVWHNRYPGARVDSETPFYQLN 73
DB 11 DVIVGAGISGVSAAWHLQDRCPNKSYYVILEKRAGMGGTWDLFRYPGIRSDSD--MYTLG 68
QY 74 IPEVWKDWTWSCRYPDQKELLSYVHHCKIRGLRKDVYFGAEVVDARYARDLGTWTVK-- 131
DB 69 FR--FRPWTCKQAIADGKPILEYIKSTAVMHGIDKYIRLNHKVTGADWSSIENRWTVQVE 126
QY 132 -TSAGHVATAKYLILATGLLH--RKHTPALPGLADFNKGKVIHSSAWHEDFDAEGQRVAVI 188
DB 127 NNGTPRMISCSFLCSCGYNYEQGYAPTFLGSEDFTGPIIHPQHWNPEDLDVAAKNIVVI 186
QY 189 GAGATSIQIVQELAKK-ADQVTFMRRRPSYCL--PMRQRTMDRNEQTANKAYPTLFEAS 245
DB 187 GSGATAVTLVPALANSKAKHVTMLQRSPTYIVSQPAKDKI-----AA 228
QY 246 RKSRI GFPVQAPSVGIFEVSPEQREAYFEELWE---RGAENFLACQYREVMDKKANRLV 302
DB 229 RLNR-WLP-----DKYAYTAVRWKNILQSALYGACQK----- 260
QY 303 YDFWAKKTRSRIVNPAKRDLMAPLEPPYWFGTKRSPLE-----SDYYEMLDKPSVEI 354
DB 261 ---WPQMRNILLGHVARRLPKGYDVQKHFGPHYNPWEQRLCLVPDGLFRAIRKGTVD 317
QY 355 VNLEQSPIVAVTKTGVLSDGSKRECDTIVLATGFD--SFTGSLTHMGLKKNKHGVDLKEV 412
DB 318 VT---DAIDRFTSTGIRLKSGBNELRADIIIVTATGLNLQFGGAVATV---DGQPVDLAQ 371
QY 413 WKDGISTYMGVFSHGFPNFAFF-VATAQAAPTVLNSNGPTIETQVDLIADTIAKL 464
DB 372 ----MSYKGMMLSGLPNMIYTCYTNASWTL-----KADLVSEFFCRL 410
RESULT 13
B83493
conserved hypothetical protein PA1208 [imported] - *Pseudomonas aeruginosa* (strain P
C;Species: *Pseudomonas aeruginosa*
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: B83493
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.;
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic
A;Reference number: A82950; MUID:20437337
A;Accession: B83493
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-499 <STO>
A;Cross-references: GB:AE004551; GB:AE004091; NID:g9947135; PIDN:AAG04597.1; GSPDB
A;Experimental source: strain PA01
C;Genetics:
A;Gene: PA1208
Query Match 11.8%; Score 342.5; DB 2; Length 499;
Best Local Similarity 26.0%; Pred. No. 7.5e-19;
Matches 114; Conservative 53; Mismatches 183; Indels 89; Gaps 13;
QY 16 DAIIVGAGLSGISAVYKLRKRLNAKIF---EGAPDFGGVWHNRYPGARVDSETPFYQL 72
DB 7 DVLIVGAGLSGVGAAYHLMK-HCPCKSFALLEGRAAMGGTWDLFRYPGIRSDSD----MF 61

Db 257 -----PEKVKERLLGMV-RDHLGPGYDVETHFTPRYNPDQRLCLVDPDADLFDAIKSGAA 310
QY 353 EIVNLEQSPIVAVTKTGVLLSDGSKRECDTIVLATGFDSTGSLTHMGLKNKHGVDLKEV 412
Db 311 SVVT---DHIEFTETGTGQLKSGKTLADADVIVTATG-----LQLQLLSGM---EV 354
QY 413 WKDG-----ISTYMGVFSHGFPN 430
Db 355 VVDGKVADLSQSMSYKGMFMFSDVPN 379

Search completed: April 27, 2002, 07:48:20
Job time: 6727 sec

Db 303 VKEITKGAQKMDQEKDFDSIIIFATGYKSNVPTWLOGGDFFTDDGMPK---TPFPNGWR 359

QY 415 DGISTYMGVESHGFPNAFFVATA 437

Db 360 GG---KGLYTVGFTRRGLLGT 378

RESULT 2

PCT-US02-09921-814

; Sequence 814, Application PC/TUS0209921

; GENERAL INFORMATION:

; APPLICANT: INCYTE GENOMICS, INC.

; APPLICANT: DAFFO, Abel

; APPLICANT: JONES, Anissa L.

; APPLICANT: TRAN, Alanna-Phung B.

; APPLICANT: DAHL, Christopher R.

; APPLICANT: GIETZEN, Darryl

; APPLICANT: CHINN, Joyce

; APPLICANT: DUFOUR, Gerard E.

; APPLICANT: HILLMAN, Jennifer L.

; APPLICANT: YU, Jimmy Y.

; APPLICANT: TUASON, Olivia

; APPLICANT: YAP, Pierre E.

; APPLICANT: AMSHEY, Stefan R.

; APPLICANT: DAUGHERTY, Sean C.

; APPLICANT: DAM, Tam C.

; APPLICANT: LIU, Tommy F.

; APPLICANT: NGUYEN, Duy-Viet An

; APPLICANT: KLEEFELD, Yael

; APPLICANT: GERSTIN JR., Edward H.

; APPLICANT: PERALTA, Careyna H.

; APPLICANT: DAVID, Marie H.

; APPLICANT: LEWIS, Samantha A.

; APPLICANT: CHEN, Alice J.

; APPLICANT: PANZER, Scott R.

; APPLICANT: HARRIS, Bernard

; APPLICANT: FLORES, Vincent

; APPLICANT: MARWAHA, Rakesh

; APPLICANT: LO, Audrey

; APPLICANT: LAN, Ruth Y.

; APPLICANT: URASHKA, Michael

; TITLE OF INVENTION: SECRETORY MOLECULES

; FILE REFERENCE: PT-1232 PCT

; CURRENT APPLICATION NUMBER: PCT/US02/09921

; CURRENT FILING DATE: 2002-03-27

; PRIOR APPLICATION NUMBER: 60/280,067; 60/280,068; 60/291,280; 60/291,849;

; 60/291,829; 60/299,428; 60/300,001; 60/299,776

; PRIOR FILING DATE: 2001-03-29; 2001-03-29; 2001-05-16; 2001-05-17;

; 2001-05-17; 2001-06-19; 2001-06-20; 2001-06-20

; NUMBER OF SEQ ID NOS: 1146

; SOFTWARE: PERL Program

; SEQ ID NO 814

; LENGTH: 717

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc_feature

; OTHER INFORMATION: Incyte ID NO: LI:1018424.4.orf2:2001MAY17

PCT-US02-09921-814

Query Match 4.3%; Score 125.5; DB 1; Length 717;

Best Local Similarity 23.7%; Pred. No. 0.0025;

Matches 59; Conservative 36; Mismatches 103; Indels 51; Gaps 12;

QY 16 DAIIVGAGLSGISAVYKLRKLRNAKIFEGAPDFGGVWHNNRYPGAR-VDSETPFY---- 70

Db 65 DYCVLGAGPAGLQAMAYFLQAGRDAYVFERAPRGS--FFTRYPRHRKLLISINKRYTGKA 122

QY 71 --QLNIPEVWKDWT-----WSCRY-PDQKELLSYVHHCDKIRGLKRDVYFGAEV 117

Db 123 NAEFNLRHWDNSLLSHDPRLLFRHYSRAYFPDARDMVRYLGDFADTLGLR--VQYNTTIA 180

QY 118 DARYARDLGTWT-----VKTSAGHVATAKYLILATGLLHRKHTPALPGLADFNKGVI--- 169

Db 181 HVTLDKDRQAWNGHYFILTQKGVHQC SVLFVATGL-----SVPNQVDFPGSEYAE 233

QY 170 HSSAWHEDFDAEGQRVAVIGAGATSIQIVQELAKKADQVTFMFMRRPSYCLPMRQRTMDRN 229

Db 234 YESVSDPEDFVGQNVLIILGRNSAF-----ETAENILGVTFNFIHMLS-----RSRV----- 280

QY 230 EOTAWKAYY 238

Db 281 -RLSWATHY 288

RESULT 3

US-09-540-209B-8514

; Sequence 8514, Application US/09540209B

; GENERAL INFORMATION:

; APPLICANT: Gary L. Breton

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTERIOIDES

; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 2709.1001-001

; CURRENT APPLICATION NUMBER: US/09/540,209B

; CURRENT FILING DATE: 2000-04-04

; NUMBER OF SEQ ID NOS: 10444

; SEQ ID NO 8514

; LENGTH: 320

; TYPE: PRT

; ORGANISM: B.fragilis

US-09-540-209B-8514

Query Match 3.7%; Score 106; DB 5; Length 320;

Best Local Similarity 24.7%; Pred. No. 0.04;

Matches 61; Conservative 33; Mismatches 79; Indels 74; Gaps 12;

QY 18 IIVGAGLSGISAVYKLRKLRNAKIFEGAPDFGGVWHNNRYPGARVDSET-----PEYQL 72

Db 14 LIIGSGPAGYTAAYAGRANLSPVLYEGI-----OPGGQLTTTDDVENFPGYPQ 62

QY 73 NI--PEVWKDWTWSCRYPDQKELLSYVHHCDKIRGLRKDV-YFGAEVVDAKY-----ARDL 125

Db 63 GISGPQLMED-----LRTQAERFGA---DIRFGIATASDL 94

QY 126 G--TWTVKTSAGHVATAKYLILATGLLHRKHTPALPGLAD---FNGKVIHSSAWHEDFDA 180

Db 95 GQAPYKITIDGEKVEADSLIIATGA-----TAKYLGLDDEKKYAGMGVSACATCDGFFY 149

QY 181 EGQRVAVIGAGATSIQIVQELAKKADQVTFMFMRRP-----SYCLPMRQRTMDRNEQTAWKA 236

Db 150 RKKVAVVWGGGTACEAIYLAGLASKVYLVVRKPYLRASKIMQERVVRKHKDIE----- 203

QY 237 YYPTLFE 243

Db 204 ---VLFE 207

RESULT 4

US-09-540-209B-10395

; Sequence 10395, Application US/09540209B

; GENERAL INFORMATION:

; APPLICANT: Gary L. Breton

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTERIOIDES

; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 2709.1001-001

; CURRENT APPLICATION NUMBER: US/09/540,209B

; CURRENT FILING DATE: 2000-04-04

; NUMBER OF SEQ ID NOS: 10444

; SEQ ID NO 10395

; LENGTH: 491

; TYPE: PRT

; ORGANISM: B.fragilis

; FEATURE:

; NAME/KEY: UNSURE

QY	104	RGLRKDVYFGAEVVD--ARYARDLGTWTVKTSAGH-----VATAKYLI-LATGILLHRK	1533
Db	432	KGKEKDPLAALVRDILSRTEKGVGSKLPGKGGHGGFVLSVILANHAIKLLTSLFQDL	4913
QY	154	HTPALPLADFNKGKVIHSSAWHEDEDAEQQRVAV-IGAGATSIQIVQELAKKADQVTNFM	2123
Db	492	QVEAL-----HKGWETDGPAAALSIMAQSTSIQRIQRLIDSVPLMNLLL	5353

QY	17	AIIVG----	AGLSGISAVYKLRKLRLNAKIFEGAPDFGGVHHNNRYPGARVDSETP-FY	70
Db	137	AVLLGDQKAVKGGSG----	KVIASKPNDFHIFVYADHGG-----	PVGLGMPNTPHIY 184
QY	71	QLNIPEVWKDWTWSCRYPDQKELLSYVHHCDKIRGLRKDVYFGAEVVDARYARDLGTWTV	1300	
Db	185	AADFIEITLKKKHASGTY--KEMVIYVEACE-----	SGSIFEGIMPKDLNIYV- 2299	
QY	131	KTSAGHVATAKYLILATGLLHRKHTPALPGLADFNKGKVIHSSAWHEDFDAEQORVAVIGA	1900	
Db	230	-TTASNAQESSYGTGCPGMNPPSPSEYITCLGD----	LYSVAMMEDSETHNLKKEI-- 2811	
QY	191	GATSIQIVQELAKKADQVTMFEMRRPSY-----	CLPMRQRTMDRNEQTAWKAYVPT-- 2400	
Db	282	-----KQYHTVKMRTSNYNTYSGGSHVMEYGNNSIKSEKLYLYQGFDPATV	3288	
QY	241	---LFEASRKSRIGFPVQAPSVGIFEVSPBQREAYFEELWERGAFNLACQYREVMVDKK	2977	
Db	329	NLPLNELPVKSKIGV-----	VNQRDADLLFLWH-----MYRTSEDCSR 3666	
QY	298	ANRLVYDFWAKKTRSRIVNPAKRDMLAPLEPPYWFGTKRSPLES DYEMLDKPSVEIVNL	3570	

Db 367 KKDDTLKELTETTRHRKHLDAVELIATI-----LFG-----PTMNVNLN 406
QY 358 EQSPIVAVTKTGVLSDGSKRECDTIVLATGFDSTGSLTHMGLKN-----KHGVDL 409
Db 407 VREP-----GLPLVD--DWECLKSMVRV-FEEHCGSLTQYGMKHMRAFANVCNNGVS- 455
QY 410 KEVKWDGISTYMGVFS 425
Db 456 KEMEERASTAACGGYS 471

RESULT 7
US-10-106-698-6386
; Sequence 6386, Application US/10106698
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 6386
; LENGTH: 539
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (2)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (7)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (8)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (11)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
US-10-106-698-6386

Query Match 2.9%; Score 85; DB 6; Length 539;
Best Local Similarity 19.7%; Pred. No. 5.5;
Matches 102; Conservative 49; Mismatches 175; Indels 192; Gaps 23;
QY 67 TPFYQLNIPEVWKDWTWSCRYPDOKELLSYVHHCDKIRGLRKDVYFGAEVVDARYADLG 126
Db 14 TPSF--NTTHYRESW-YACRYRSRGIPGSGTHASASTTVPGLSEE----- 53
QY 127 TWTVKTSAGHVATAKYLILATGLLHRKHTPALPGLADFNKVI-----HSSAWH 175
Db 54 STTFYSSPGSTETTAFHSNTMSIHSQQSTPPFDPSPGFTHTVLPATLTITDTCQESTAFH 113
QY 176 EDFDAEGORVAVIGAGATSIQIVQELAKKADQVTFMFMRRPSYCLPMRQRTMDRNEQTAWK 235
Db 114 SSSDATG--TTPLPARSTASDLVGE-----PTTFYISPS----- 145
QY 236 AYYPTLFEASRKSRIQIVQELAKKADQVTFMFMRRPSYCLPMRQRTMDRNEQTAWK 295
Db 146 PTYTTLFPAS--SSTSGLTEESTT---FHTSp----- 172
QY 296 KKANRLVYDFWAKKTRSRIVNPAKRDLMAP---LEPPYWFGTK-----RSPL 340
Db 173 -----SFTSTIVSTESLETAPGLCQEQIWNKGQVCPCQYVGYQCLSPLE 219

QY 341 SDYYEMLDKPSVEIVNLEQSPIVAVTKTGVL--LSDGSKRECDTIVLATGFDSTGSLTH 398
Db 220 SFPVETPEK-----LNATLGMTVKVYRNFETKMNDAASSQE-----YQNFSTL--- 262
QY 399 MGLKNKHGVDLKEVWKDGISTYMGVFSHGFPNAPFAVATAQAAPTIVLSNGPTIETQVDLIA 458
Db 263 --FKNRMDVWLK---GDNLPQYRGVNIRR-----LLNGSIVVKNDVILEA 302
QY 459 DTIAKLEAEHATSVEATKSAQEAWSIMIAKMNEHTLFLPLTDSWWTGNGNIPGKATRAL--- 515
Db 303 DYTLEYEELFENLAEIVK-----AKIMNETRTTLLD-----PDSCRKAILCY 344
QY 516 ----TFIG-----GIALYEQICQEKVANWDGF---DVL 541
Db 345 SEEDTFVDSSVTPGDFQEQCTQKAAEGYTQFYVVDVL 382

RESULT 8
US-09-573-655B-1983
; Sequence 1983, Application US/09573655B
; GENERAL INFORMATION:
; APPLICANT: SOLOVYEV, Victor and TROUKHAN, Maxim
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypept
; TITLE OF INVENTION: Thereby
; FILE REFERENCE: 2750-0876P
; CURRENT APPLICATION NUMBER: US/09/573,655B
; CURRENT FILING DATE: 2000-05-18
; NUMBER OF SEQ ID NOS: 3281
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1983
; LENGTH: 1156
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-573-655B-1983

Query Match 2.9%; Score 83.5; DB 5; Length 1156;
Best Local Similarity 20.1%; Pred. No. 22;
Matches 102; Conservative 66; Mismatches 175; Indels 165; Gaps 24;
QY 123 RDLGTWTVK--TSAGHVATAKYL-----ILATGLLHRKH---TPALPGLADFNKGV 168
Db 19 RDFGNYGKLESTMSSVSIGKRIFEGCLVLGGIIRWGVKKEARVAISPATVMSGQFN--- 75
QY 169 IHSSAWHEDFDAEGQRVAVIGAGATSIQIVQELA-----KKADQVTFM 211
Db 76 --NSPRGEDKVE-----AGTSFTEYEDSPFDIASTKNAPVERLRRWRQAALVUNA 125
QY 212 MRRPSYCLPMRQ-----RTMDRNEQTAWKAYYPTLFE--ASRKSRIQIVQAPSVEI 261
Db 126 SRRFRYTDLKREEDKKQMLRKMRAHAQAIRAAH---LFKAAASRVGTIASPLTPGGGD 182
QY 262 FEVSPEQ-----REAYFEELWERGAENFLACQYR-----EVMVDKKANRLVYDF 305
Db 183 FGIGQEQIVSISRDQNIQALQELGVRGLSDLLKTNLEKGIHGDDDDILKRKSAFGSNTY 242
QY 306 WAKKTRS--RIVNPAKRDL-----MAPLEPPYWFGTKRSPLESYYEMLOKPSVEI 354
Db 243 PQKKGRSFRWFVEASODLTLLIIVAASLASLA---LGIKTEGIEKGWYDGIS----- 292
QY 355 VNLEQSPIVAVTKTGVLSDGSKRECDTIVLATGFDSTGSLTHMGL-KNKHGVDLKEVW 413
Db 293 -----IAFAVLL-----VIVVTATSDYRQSLQFQNLNEEKRNIRL-EVT 330
QY 414 KDG----ISTY-----MGVFSHGFPNAPFAFFVATAQAAPTIVLSNGPTIET 452
Db 331 RDGRRVEISYDIVGDVIPLENGDQASFMLSMQYFLALFRISVPADGVLVACHSLAVD 390
QY 453 QVDLIADT--IAKLEAEHATSVEATKSA-----QEAWSIMIAKMNEHTLFP 497
Db 391 ESSMTGESKIVQKNSTKHPFLMSGCKVADGNGTMLVTGVGVNTEWGLLMASVSED----- 445
QY 498 TDSWWTGGNIP--GKATRALTFIGGIAL 523


```

; ORGANISM: Arabidopsis thaliana
US-09-573-655B-117

Query Match      2.8%; Score 81; DB 5; Length 947;
Best Local Similarity 23.2%; Pred. No. 27;
Matches 94; Conservative 47; Mismatches 157; Indels 108; Gaps 23;

Qy 152 RKHTPALPLGLADF---NGKVIHSS-----AWHEFED-AEQQRVAVIGAGATSIQIVQE 200
|| || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 212 RKATPKGLKLRKRIEAPDGTGLVHDSSYVGENAWDDDDLETTGSLKKIIGNAR-----IQT 267

Qy 201 LAKK-----ADQVTMFMRPSPSYCLPMRQRTMDRNEQTAWKAYYPTLFEASRKS 248
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 268 EAKKLSQDLGVSGEIGDSVGNWRERLATWKEMLEREKLSEQLNSSAAKYVVEFDMKEVE 327

Qy 249 RIGFPVQAPSVGIFEVSPEQREAYFEELWERGAFNACQYREVVMVDKKANRLVYDFWAK 308
: : : : : : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 328 K---SLREDVIGRTSETEGTRALWISKRWWR-----YRP-----KLPYTYFLQ 367

Qy 309 KTRSRIV-----NPAKRDLMA-----PLEPPYWFGTKRSPLESYDEMIDKPSVEIVNL 357
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 368 KLDSSVAAVVFTEDLKRLLYVTMKEGPPLE--YIVDI---PLDPYLEETICNAGVEVDLL 422

Qy 358 EQSPIVAVTKTGVLLSDGSK-----REC DTIVLATG-----PD-----SF---T 393
: : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 423 QKRQIHVFMKVFIALLPGILILWFIFRESAMLLLITSKRFLYKKYNQLFDMAYAENFILPV 482

Qy 394 GSLTHMGLKNKHGVDLKEVWK--DGISTYMGVFSHGFPNAFF---VATAQAAPT VLSNGPT 448
| : : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 483 GDVSETKSMYKEWVLGGDVWDLLDELMYM-----GNPMQYYEKDVAPVRG--VLLSGPP 535

Qy 449 ILETQVDLIADTIAK--LEAEHATSVKATKSAQEAWSIMIAKMNE 491
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 536 --GTGKTLFARTLAKESGLPFVFGAGAEFTDSEKSG---AAKINE 575
```

Search completed: April 27, 2002, 07:58:35
Job time: 736 sec

```

RESULT 15
US-09-540-209B-6436
; Sequence 6436, Application US/09540209B
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES FR
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1001-001
; CURRENT APPLICATION NUMBER: US/09/540,209B
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 6436
; LENGTH: 327
; TYPE: PRT
; ORGANISM: B.fragilis
US-09-540-209B-6436

Query Match      2.8%; Score 80.5; DB 5; Length 327;
Best Local Similarity 20.7%; Pred. No. 6.7;
Matches 67; Conservative 38; Mismatches 112; Indels 107; Gaps 17;

Qy 255 QAPSVGIFEVSPEQR--EAYFEELWERGAFNACQY-----REVMDKKANRL 301
| : | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 55 QSPMTGVLEAYPCQAGGAIYYGNYIEGKLTSPFGMYLQNGEIIYGDKNREISLPVGTYNM 114

Qy 302 VYDFWAKKTRSRIV--NPAKRDLMAP-----LEPPYWFGTKRSPLESYDEMID---- 348
: | | | | : | | : | | | | | | | | | | | | | | | | | | | | | | |
Db 115 IY--WGTPKYEELIYSNPV--VVAPQITIGGDLSSQY-FGLRKVSADTTYYPVFDLVYT 168

Qy 349 -KPS---VEIVNLEQSPIVATK-----TGVLSDGSKRECDTIVLATGDSFTGSLT 397
|| : | : : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 169 VKPAHIGTEELSAAMQQRVAVAGLKVIYKNKNNGILSS-----SIAGMEV 211

Qy 398 HMGLKNKHGVDLKEVWKDGISTYMGVFSHGFPNAFFVATAQAAPT VLSNGPTIETQVDLI 457
| : | | : : : : | | | | | | | | | | | | | | | | | | | | | | | |
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2002, 07:45:39 ; Search time 389.72 Seconds
(without alignments)
492.224 Million cell updates/sec

Title: US-09-882-694A-3
Perfect score: 2902
Sequence: 1 MSATNSNRGDCSVACDAIIV.....QICQEKVANWDGFDVLHAPC 545

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 3502263 seqs, 351980561 residues
Total number of hits satisfying chosen parameters: 3502263

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_AA_Main:
1: /cgn2_6/ptodata/2/paa/PCTUS_COMB.pep.*
2: /cgn2_6/ptodata/2/paa/US06_COMB.pep.*
3: /cgn2_6/ptodata/2/paa/US07_COMB.pep.*
4: /cgn2_6/ptodata/2/paa/US080_COMB.pep.*
5: /cgn2_6/ptodata/2/paa/US081_COMB.pep.*
6: /cgn2_6/ptodata/2/paa/US082_COMB.pep.*
7: /cgn2_6/ptodata/2/paa/US083_COMB.pep.*
8: /cgn2_6/ptodata/2/paa/US084_COMB.pep.*
9: /cgn2_6/ptodata/2/paa/US085_COMB.pep.*
10: /cgn2_6/ptodata/2/paa/US086_COMB.pep.*
11: /cgn2_6/ptodata/2/paa/US087_COMB.pep.*
12: /cgn2_6/ptodata/2/paa/US088_COMB.pep.*
13: /cgn2_6/ptodata/2/paa/US089_COMB.pep.*
14: /cgn2_6/ptodata/2/paa/US090_COMB.pep.*
15: /cgn2_6/ptodata/2/paa/US091_COMB.pep.*
16: /cgn2_6/ptodata/2/paa/US092_COMB.pep.*
17: /cgn2_6/ptodata/2/paa/US093_COMB.pep.*
18: /cgn2_6/ptodata/2/paa/US094_COMB.pep.*
19: /cgn2_6/ptodata/2/paa/US095_COMB.pep.*
20: /cgn2_6/ptodata/2/paa/US096_COMB.pep.*
21: /cgn2_6/ptodata/2/paa/US097_COMB.pep.*
22: /cgn2_6/ptodata/2/paa/US098_COMB.pep.*
23: /cgn2_6/ptodata/2/paa/US099_COMB.pep.*
24: /cgn2_6/ptodata/2/paa/US100_COMB.pep.*
25: /cgn2_6/ptodata/2/paa/US101_COMB.pep.*
26: /cgn2_6/ptodata/2/paa/US60_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2902	100.0	545	17 US-09-351-224-3	Sequence 3, Appli
2	2902	100.0	545	17 US-09-351-224E-3	Sequence 3, Appli
3	2902	100.0	545	17 US-09-351-823-3	Sequence 3, Appli
4	2902	100.0	545	20 US-09-677-488-3	Sequence 3, Appli
5	2902	100.0	545	20 US-09-677-488A-3	Sequence 3, Appli
6	2902	100.0	545	20 US-09-677-682-3	Sequence 3, Appli
7	2902	100.0	545	20 US-09-677-682A-3	Sequence 3, Appli

8	2902	100.0	545	20	US-09-677-682B-3	Sequence 3, Appli
9	2902	100.0	545	22	US-09-882-694-3	Sequence 3, Appli
10	2902	100.0	545	22	US-09-882-694A-3	Sequence 3, Appli
11	833.5	28.7	529	19	US-09-504-358-22	Sequence 22, Appli
12	833.5	28.7	529	23	US-09-954-314-22	Sequence 22, Appli
13	833.5	28.7	529	26	US-60-120-702-22	Sequence 22, Appli
14	804	27.7	547	20	US-09-648-004-10	Sequence 10, Appli
15	796	27.4	553	19	US-09-504-358-6	Sequence 6, Appli
16	796	27.4	553	23	US-09-954-314-6	Sequence 6, Appli
17	796	27.4	553	26	US-60-120-702-6	Sequence 6, Appli
18	690	23.8	439	18	US-09-417-507-40604	Sequence 40604, A
19	588.5	20.3	439	18	US-09-417-507-26928	Sequence 26928, A
20	546.5	18.8	522	17	US-09-328-352-6840	Sequence 6840, Ap
21	508	17.5	243	18	US-09-417-507-40608	Sequence 40608, A
22	485.5	16.7	306	18	US-09-417-507-27508	Sequence 27508, A
23	481.5	16.6	603	21	US-09-732-615-2	Sequence 2, Appli
24	481.5	16.6	603	26	US-60-170-214-2	Sequence 2, Appli
25	477.5	16.5	519	17	US-09-328-352-7853	Sequence 7853, Ap
26	465.5	16.0	538	16	US-09-252-991A-23068	Sequence 23068, A
27	453	15.6	525	17	US-09-328-352-6683	Sequence 6683, Ap
28	444	15.3	516	17	US-09-328-352-6474	Sequence 6474, Ap
29	436	15.0	528	23	US-09-902-540-12229	Sequence 12229, A
30	398	13.7	219	20	US-09-675-784A-8877	Sequence 8877, Ap
31	396	13.6	530	17	US-09-328-352-7333	Sequence 7333, Ap
32	362.5	12.5	489	22	US-09-888-320-2	Sequence 2, Appli
33	354.5	12.2	482	22	US-09-897-516-6981	Sequence 6981, Ap
34	354.5	12.2	482	26	US-60-215-161-6981	Sequence 6981, Ap
35	346.5	11.9	543	16	US-09-252-991A-19684	Sequence 19684, A
36	323	11.1	351	16	US-09-252-991A-25097	Sequence 25097, A
37	300.5	10.4	507	17	US-09-328-352-4291	Sequence 4291, Ap
38	290	10.0	253	18	US-09-417-507-44096	Sequence 44096, A
39	280	9.6	312	17	US-09-417-507-24881	Sequence 24881, A
40	262.5	9.0	558	17	US-09-351-150A-7	Sequence 7, Appli
41	249	8.6	343	16	US-09-248-796-20883	Sequence 20883, A
42	249	8.6	343	26	US-60-096-409-20883	Sequence 20883, A
43	247	8.5	335	18	US-09-417-507-41815	Sequence 41815, A
44	244.5	8.4	364	16	US-09-248-796-16427	Sequence 16427, A
45	244.5	8.4	364	26	US-60-096-409-16427	Sequence 16427, A

ALIGNMENTS

RESULT 1
US-09-351-224-3
; Sequence 3, Application US/09351224
; GENERAL INFORMATION:
; APPLICANT: Duwick, Jon
; TITLE OF INVENTION: Compositions and Methods for Fummonisin Detoxification
; FILE REFERENCE: 5718-111
; CURRENT APPLICATION NUMBER: US/09/351,224
; CURRENT FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 545
; TYPE: PRT
; ORGANISM: Exophiala spinifera
; FEATURE:
; OTHER INFORMATION: flavin monooxygenase
US-09-351-224-3

Query Match	100.0%	Score	2902;	DB	17;	Length	545;
Best Local Similarity	100.0%	Pred. No.	1.5e-300;				
Matches	545;	Conservative	0;	Mismatches	0;	Indels	0;
QY	1	MSATNSNRGDCSVACDAIIVGAGLSGISAVYKLRKLRNNAKIFEGAPDFGGVHWNRYPG	60				
Db	1	MSATNSNRGDCSVACDAIIVGAGLSGISAVYKLRKLRNNAKIFEGAPDFGGVHWNRYPG	60				
QY	61	ARVDSETPFYQLNIPEVWKDWTWSCRYPDQKELLSYVHHCDKIRGLRKDVYFGAEVVDAR	120				

```
Db 61 ARVDSETPFYQLNIPEVWKDWTWSCRYPDQKELLSYVHHCDKIRGLRKDVYFGAEVVDAR 120
QY 121 YARDLGTWTVKTSAGHVATAKYILILATGLLHRKHHTPALPGLADFNKGKVIHSSAWHEDFDA 180
Db 121 YARDLGTWTVKTSAGHVATAKYILILATGLLHRKHHTPALPGLADFNKGKVIHSSAWHEDFDA 180
QY 181 EGQRVAVIGAGATSIOIVQELAKKADQVTFMFMRRPSYCLPMRQRTMDRNEQTAWKAYYPT 240
Db 181 EGQRVAVIGAGATSIOIVQELAKKADQVTFMFMRRPSYCLPMRQRTMDRNEQTAWKAYYPT 240
QY 241 LFEASRKSRIIGFPVQAPSVGIFEVSPEQREAYFEELWERGAPNFLACQYREVMVDKKNR 300
Db 241 LFEASRKSRIIGFPVQAPSVGIFEVSPEQREAYFEELWERGAPNFLACQYREVMVDKKNR 300
QY 301 LVYDFWAKKTRRSRIVNPAKRDLMAPEPPYWFGTKRSPLESDDYEMLDKPSVEIVNLEQS 360
Db 301 LVYDFWAKKTRRSRIVNPAKRDLMAPEPPYWFGTKRSPLESDDYEMLDKPSVEIVNLEQS 360
QY 361 PIVAVTKTGVLSDGSKRECDTIVLATGDFSFTGSLTHMGLKNKHGVLDKEVWKDGISTY 420
Db 361 PIVAVTKTGVLSDGSKRECDTIVLATGDFSFTGSLTHMGLKNKHGVLDKEVWKDGISTY 420
QY 421 MGVSFHGFPNAFFVATAQAFTVLSNGPTIETQVDLIADTIAKLEAEHATSVEATKSAQE 480
Db 421 MGVSFHGFPNAFFVATAQAFTVLSNGPTIETQVDLIADTIAKLEAEHATSVEATKSAQE 480
QY 481 AWSIMIAKMNEHTLFPLTDSWWTGGNIPGKATRALTFIGGIALYEQICQEKVANWDGFDV 540
Db 481 AWSIMIAKMNEHTLFPLTDSWWTGGNIPGKATRALTFIGGIALYEQICQEKVANWDGFDV 540
QY 541 LHAPC 545
Db 541 LHAPC 545
```

```
RESULT 2
US-09-351-224E-3
; Sequence 3, Application US/09351224E
; GENERAL INFORMATION:
; APPLICANT: Duivick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; TITLE OF INVENTION: Detoxification
; FILE REFERENCE: 5718-111
; CURRENT APPLICATION NUMBER: US/09/351,224E
; CURRENT FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 545
; TYPE: PRT
; ORGANISM: Exophiala spinifera
US-09-351-224E-3
```

```
Query Match 100.0%; Score 2902; DB 17; Length 545;
Best Local Similarity 100.0%; Pred. No. 1.5e-300;
Matches 545; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSATSNRSGDCSVACDAIIVGAGLSGISAIVYKLRKRLNAKIFEGAPDFGGVHHNRYPG 60
Db 1 MSATSNRSGDCSVACDAIIVGAGLSGISAIVYKLRKRLNAKIFEGAPDFGGVHHNRYPG 60

QY 61 ARVDSETPFYQLNIPEVWKDWTWSCRYPDQKELLSYVHHCDKIRGLRKDVYFGAEVVDAR 120
Db 61 ARVDSETPFYQLNIPEVWKDWTWSCRYPDQKELLSYVHHCDKIRGLRKDVYFGAEVVDAR 120

QY 121 YARDLGTWTVKTSAGHVATAKYILILATGLLHRKHHTPALPGLADFNKGKVIHSSAWHEDFDA 180
Db 121 YARDLGTWTVKTSAGHVATAKYILILATGLLHRKHHTPALPGLADFNKGKVIHSSAWHEDFDA 180
```

```
QY 181 EGQRVAVIGAGATSIOIVQELAKKADQVTFMFMRRPSYCLPMRQRTMDRNEQTAWKAYYPT 240
Db 181 EGQRVAVIGAGATSIOIVQELAKKADQVTFMFMRRPSYCLPMRQRTMDRNEQTAWKAYYPT 240
QY 241 LFEASRKSRIIGFPVQAPSVGIFEVSPEQREAYFEELWERGAPNFLACQYREVMVDKKNR 300
Db 241 LFEASRKSRIIGFPVQAPSVGIFEVSPEQREAYFEELWERGAPNFLACQYREVMVDKKNR 300
QY 301 LVYDFWAKKTRRSRIVNPAKRDLMAPEPPYWFGTKRSPLESDDYEMLDKPSVEIVNLEQS 360
Db 301 LVYDFWAKKTRRSRIVNPAKRDLMAPEPPYWFGTKRSPLESDDYEMLDKPSVEIVNLEQS 360
QY 361 PIVAVTKTGVLSDGSKRECDTIVLATGDFSFTGSLTHMGLKNKHGVLDKEVWKDGISTY 420
Db 361 PIVAVTKTGVLSDGSKRECDTIVLATGDFSFTGSLTHMGLKNKHGVLDKEVWKDGISTY 420
QY 421 MGVSFHGFPNAFFVATAQAFTVLSNGPTIETQVDLIADTIAKLEAEHATSVEATKSAQE 480
Db 421 MGVSFHGFPNAFFVATAQAFTVLSNGPTIETQVDLIADTIAKLEAEHATSVEATKSAQE 480
QY 481 AWSIMIAKMNEHTLFPLTDSWWTGGNIPGKATRALTFIGGIALYEQICQEKVANWDGFDV 540
Db 481 AWSIMIAKMNEHTLFPLTDSWWTGGNIPGKATRALTFIGGIALYEQICQEKVANWDGFDV 540
QY 541 LHAPC 545
Db 541 LHAPC 545
```

```
RESULT 3
US-09-351-823-3
; Sequence 3, Application US/09351823
; GENERAL INFORMATION:
; APPLICANT: Duivick, Jon
; TITLE OF INVENTION: Compositions and Methods for Fumonisin Detoxification
; FILE REFERENCE: 5718-56
; CURRENT APPLICATION NUMBER: US/09/351,823
; CURRENT FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 3
; LENGTH: 545
; TYPE: PRT
; ORGANISM: Exophiala spinifera
; FEATURE:
; OTHER INFORMATION: flavin monooxygenase
US-09-351-823-3
```

```
Query Match 100.0%; Score 2902; DB 17; Length 545;
Best Local Similarity 100.0%; Pred. No. 1.5e-300;
Matches 545; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSATSNRSGDCSVACDAIIVGAGLSGISAIVYKLRKRLNAKIFEGAPDFGGVHHNRYPG 60
Db 1 MSATSNRSGDCSVACDAIIVGAGLSGISAIVYKLRKRLNAKIFEGAPDFGGVHHNRYPG 60

QY 61 ARVDSETPFYQLNIPEVWKDWTWSCRYPDQKELLSYVHHCDKIRGLRKDVYFGAEVVDAR 120
Db 61 ARVDSETPFYQLNIPEVWKDWTWSCRYPDQKELLSYVHHCDKIRGLRKDVYFGAEVVDAR 120

QY 121 YARDLGTWTVKTSAGHVATAKYILILATGLLHRKHHTPALPGLADFNKGKVIHSSAWHEDFDA 180
Db 121 YARDLGTWTVKTSAGHVATAKYILILATGLLHRKHHTPALPGLADFNKGKVIHSSAWHEDFDA 180

QY 181 EGQRVAVIGAGATSIOIVQELAKKADQVTFMFMRRPSYCLPMRQRTMDRNEQTAWKAYYPT 240
Db 181 EGQRVAVIGAGATSIOIVQELAKKADQVTFMFMRRPSYCLPMRQRTMDRNEQTAWKAYYPT 240

QY 241 LFEASRKSRIIGFPVQAPSVGIFEVSPEQREAYFEELWERGAPNFLACQYREVMVDKKNR 300
Db 241 LFEASRKSRIIGFPVQAPSVGIFEVSPEQREAYFEELWERGAPNFLACQYREVMVDKKNR 300
```

```
QY 301 LVYDFWAKKTRSRIVNPAKRDLMAPLEPPYWFGTKRSPLESYYEMLDKPSVEIVNLEQS 360
Db 301 LVYDFWAKKTRSRIVNPAKRDLMAPLEPPYWFGTKRSPLESYYEMLDKPSVEIVNLEQS 360
QY 361 PIVAVTKTGVLSDGSKRECDTIVLATGFDSTGSLTHMGLKNKHGVDLKEVWKDGISTY 420
Db 361 PIVAVTKTGVLSDGSKRECDTIVLATGFDSTGSLTHMGLKNKHGVDLKEVWKDGISTY 420
QY 421 MGVFSGHPNAPFFVATAQAPTIVLSNGPTIETQVDLIADTIAKLEAEHATSVEATKSAQE 480
Db 421 MGVFSGHPNAPFFVATAQAPTIVLSNGPTIETQVDLIADTIAKLEAEHATSVEATKSAQE 480
QY 481 AWSIMIAKMNEHTLFPLTDSWWTGGNIPGKATRALTFIGGIALYEQICQEKVANWDGFDV 540
Db 481 AWSIMIAKMNEHTLFPLTDSWWTGGNIPGKATRALTFIGGIALYEQICQEKVANWDGFDV 540
QY 541 LHAPC 545
Db 541 LHAPC 545

RESULT 4
US-09-677-488-3
; Sequence 3, Application US/09677488
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jon
; TITLE OF INVENTION: Compositions and Methods for Fumonisin Detoxification
; FILE REFERENCE: 35718/204100 (5718-111A )
; CURRENT APPLICATION NUMBER: US/09/677,488
; CURRENT FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 3
; LENGTH: 545
; TYPE: PRT
; ORGANISM: Exophiala spinifera
; FEATURE:
; OTHER INFORMATION: flavin monooxygenase, translation of fully spliced
; OTHER INFORMATION: cdna
US-09-677-488-3
```

```
Query Match 100.0%; Score 2902; DB 20; Length 545;
Best Local Similarity 100.0%; Pred. No. 1.5e-300;
Matches 545; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSATSNRSGDCSVACDAIIVGAGLSGISAIVYKLRKLRNNAKIFEGAPDFGCVHWNRYPG 60
Db 1 MSATSNRSGDCSVACDAIIVGAGLSGISAIVYKLRKLRNNAKIFEGAPDFGCVHWNRYPG 60
QY 61 ARVDSETPFYQLNIPEVWKDWTWSCRYPDQKELLSYVHCDKIRGLRKDVYFGAEVVDAR 120
Db 61 ARVDSETPFYQLNIPEVWKDWTWSCRYPDQKELLSYVHCDKIRGLRKDVYFGAEVVDAR 120
QY 121 YARDLGTWTVKTSAGHVATAKYLILATGLLHRKHTPALPGLADFNKGVIHSSAWHEDFDA 180
Db 121 YARDLGTWTVKTSAGHVATAKYLILATGLLHRKHTPALPGLADFNKGVIHSSAWHEDFDA 180
QY 181 EGQRVAVIGAGATSIQIVQELAKKADQVTFMRRRPSYCLPMRQRTMDRNEQTAWKAYYPT 240
Db 181 EGQRVAVIGAGATSIQIVQELAKKADQVTFMRRRPSYCLPMRQRTMDRNEQTAWKAYYPT 240
QY 241 LFEASRKSRIIGFPVQAPSVGIFEVSPEQREAYFEELWERGAFNFLACQYREVMVDKKANR 300
Db 241 LFEASRKSRIIGFPVQAPSVGIFEVSPEQREAYFEELWERGAFNFLACQYREVMVDKKANR 300
QY 301 LVYDFWAKKTRSRIVNPAKRDLMAPLEPPYWFGTKRSPLESYYEMLDKPSVEIVNLEQS 360
Db 301 LVYDFWAKKTRSRIVNPAKRDLMAPLEPPYWFGTKRSPLESYYEMLDKPSVEIVNLEQS 360
```

```
QY 361 PIVAVTKTGVLSDGSKRECDTIVLATGFDSTGSLTHMGLKNKHGVDLKEVWKDGISTY 420
Db 361 PIVAVTKTGVLSDGSKRECDTIVLATGFDSTGSLTHMGLKNKHGVDLKEVWKDGISTY 420
QY 421 MGVFSGHPNAPFFVATAQAPTIVLSNGPTIETQVDLIADTIAKLEAEHATSVEATKSAQE 480
Db 421 MGVFSGHPNAPFFVATAQAPTIVLSNGPTIETQVDLIADTIAKLEAEHATSVEATKSAQE 480
QY 481 AWSIMIAKMNEHTLFPLTDSWWTGGNIPGKATRALTFIGGIALYEQICQEKVANWDGFDV 540
Db 481 AWSIMIAKMNEHTLFPLTDSWWTGGNIPGKATRALTFIGGIALYEQICQEKVANWDGFDV 540
QY 541 LHAPC 545
Db 541 LHAPC 545

RESULT 5
US-09-677-488A-3
; Sequence 3, Application US/09677488A
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; FILE REFERENCE: 35718/204100
; CURRENT APPLICATION NUMBER: US/09/677,488A
; CURRENT FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 545
; TYPE: PRT
; ORGANISM: Exophiala spinifera
; OTHER INFORMATION:
US-09-677-488A-3
```

```
Query Match 100.0%; Score 2902; DB 20; Length 545;
Best Local Similarity 100.0%; Pred. No. 1.5e-300;
Matches 545; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSATSNRSGDCSVACDAIIVGAGLSGISAIVYKLRKLRNNAKIFEGAPDFGCVHWNRYPG 60
Db 1 MSATSNRSGDCSVACDAIIVGAGLSGISAIVYKLRKLRNNAKIFEGAPDFGCVHWNRYPG 60
QY 61 ARVDSETPFYQLNIPEVWKDWTWSCRYPDQKELLSYVHCDKIRGLRKDVYFGAEVVDAR 120
Db 61 ARVDSETPFYQLNIPEVWKDWTWSCRYPDQKELLSYVHCDKIRGLRKDVYFGAEVVDAR 120
QY 121 YARDLGTWTVKTSAGHVATAKYLILATGLLHRKHTPALPGLADFNKGVIHSSAWHEDFDA 180
Db 121 YARDLGTWTVKTSAGHVATAKYLILATGLLHRKHTPALPGLADFNKGVIHSSAWHEDFDA 180
QY 181 EGQRVAVIGAGATSIQIVQELAKKADQVTFMRRRPSYCLPMRQRTMDRNEQTAWKAYYPT 240
Db 181 EGQRVAVIGAGATSIQIVQELAKKADQVTFMRRRPSYCLPMRQRTMDRNEQTAWKAYYPT 240
QY 241 LFEASRKSRIIGFPVQAPSVGIFEVSPEQREAYFEELWERGAFNFLACQYREVMVDKKANR 300
Db 241 LFEASRKSRIIGFPVQAPSVGIFEVSPEQREAYFEELWERGAFNFLACQYREVMVDKKANR 300
QY 301 LVYDFWAKKTRSRIVNPAKRDLMAPLEPPYWFGTKRSPLESYYEMLDKPSVEIVNLEQS 360
Db 301 LVYDFWAKKTRSRIVNPAKRDLMAPLEPPYWFGTKRSPLESYYEMLDKPSVEIVNLEQS 360
QY 361 PIVAVTKTGVLSDGSKRECDTIVLATGFDSTGSLTHMGLKNKHGVDLKEVWKDGISTY 420
Db 361 PIVAVTKTGVLSDGSKRECDTIVLATGFDSTGSLTHMGLKNKHGVDLKEVWKDGISTY 420
```



```
QY 421 MGVSFHGFPNAFFVATAQAAPTIVLSNGPTIIETQVDLIADTIAKLEAEHATSVEATKSAQE 480
|||||
Db 421 MGVSFHGFPNAFFVATAQAAPTIVLSNGPTIIETQVDLIADTIAKLEAEHATSVEATKSAQE 480

QY 481 AWSIMIAKMNEHTLFPLTDSWWTGGNIPGKATRALTFIGGIALYEIQCOEKVANWDGFDV 540
|||||
Db 481 AWSIMIAKMNEHTLFPLTDSWWTGGNIPGKATRALTFIGGIALYEIQCOEKVANWDGFDV 540

QY 541 LHAPC 545
|||||
Db 541 LHAPC 545

RESULT 6
US-09-677-682-3
; Sequence 3, Application US/09677682
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jon
; TITLE OF INVENTION: Compositions and Methods for Fummonisin Detoxification
; FILE REFERENCE: 35718/204101 (5718-111B )
; CURRENT APPLICATION NUMBER: US/09/677,682
; CURRENT FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 545
; TYPE: PRT
; ORGANISM: Exophiala spinifera
; FEATURE:
; OTHER INFORMATION: flavin monooxygenase, translation of fully spliced
; OTHER INFORMATION: cDNA
US-09-677-682-3
```

```
Query Match 100.0%; Score 2902; DB 20; Length 545;
Best Local Similarity 100.0%; Pred. No. 1.5e-300;
Matches 545; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSATSNSRGDCSVACDAIIVGAGLSGISAVYKLRKLRNNAKIFEGAPDFGGVWHNRYPG 60
|||||
Db 1 MSATSNSRGDCSVACDAIIVGAGLSGISAVYKLRKLRNNAKIFEGAPDFGGVWHNRYPG 60

QY 61 ARVDSETPFYQLNIPEVWKDWTWSCRYPDQKELLSYVHCDKIRGLRKDVYFGAEVVDAR 120
|||||
Db 61 ARVDSETPFYQLNIPEVWKDWTWSCRYPDQKELLSYVHCDKIRGLRKDVYFGAEVVDAR 120

QY 121 YARDLGTWTVKTSAGHVATAKYLILATGLLHRKHHTPALPGLADFNKGKVIHSSAWHEDFDA 180
|||||
Db 121 YARDLGTWTVKTSAGHVATAKYLILATGLLHRKHHTPALPGLADFNKGKVIHSSAWHEDFDA 180

QY 181 EGQRVAVIGAGATSIQIVQELAKKADQVTMFMRPPSYCLPMRQRTMDRNEQTAWKAYYPT 240
|||||
Db 181 EGQRVAVIGAGATSIQIVQELAKKADQVTMFMRPPSYCLPMRQRTMDRNEQTAWKAYYPT 240

QY 241 LFEASRKSRIGFPVQAPSVGIFEVSPEQREAYFEELWERGAFNFLACQYREVMVDKKANR 300
|||||
Db 241 LFEASRKSRIGFPVQAPSVGIFEVSPEQREAYFEELWERGAFNFLACQYREVMVDKKANR 300

QY 301 LVYDFWAKKTRSRIVNPAKRDLMAPLEPPYWFGTKRSPLESDDYYEMLDKPSVEIVNLEQS 360
|||||
Db 301 LVYDFWAKKTRSRIVNPAKRDLMAPLEPPYWFGTKRSPLESDDYYEMLDKPSVEIVNLEQS 360

QY 361 PIVAVTKTGVLSDGSKRECDTIVLATGFDSDFTGSLTHMGLKNKKGVDLKEVWKDGISTY 420
|||||
Db 361 PIVAVTKTGVLSDGSKRECDTIVLATGFDSDFTGSLTHMGLKNKKGVDLKEVWKDGISTY 420

QY 421 MGVSFHGFPNAFFVATAQAAPTIVLSNGPTIIETQVDLIADTIAKLEAEHATSVEATKSAQE 480
|||||
Db 421 MGVSFHGFPNAFFVATAQAAPTIVLSNGPTIIETQVDLIADTIAKLEAEHATSVEATKSAQE 480
```

```
QY 481 AWSIMIAKMNEHTLFPLTDSWWTGGNIPGKATRALTFIGGIALYEIQCOEKVANWDGFDV 540
|||||
Db 481 AWSIMIAKMNEHTLFPLTDSWWTGGNIPGKATRALTFIGGIALYEIQCOEKVANWDGFDV 540

QY 541 LHAPC 545
|||||
Db 541 LHAPC 545

RESULT 7
US-09-677-682A-3
; Sequence 3, Application US/09677682A
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; TITLE OF INVENTION: Detoxification
; FILE REFERENCE: 35718/204101
; CURRENT APPLICATION NUMBER: US/09/677,682A
; CURRENT FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 545
; TYPE: PRT
; ORGANISM: Exophiala spinifera
; OTHER INFORMATION:
US-09-677-682A-3
```

```
Query Match 100.0%; Score 2902; DB 20; Length 545;
Best Local Similarity 100.0%; Pred. No. 1.5e-300;
Matches 545; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSATSNSRGDCSVACDAIIVGAGLSGISAVYKLRKLRNNAKIFEGAPDFGGVWHNRYPG 60
|||||
Db 1 MSATSNSRGDCSVACDAIIVGAGLSGISAVYKLRKLRNNAKIFEGAPDFGGVWHNRYPG 60

QY 61 ARVDSETPFYQLNIPEVWKDWTWSCRYPDQKELLSYVHCDKIRGLRKDVYFGAEVVDAR 120
|||||
Db 61 ARVDSETPFYQLNIPEVWKDWTWSCRYPDQKELLSYVHCDKIRGLRKDVYFGAEVVDAR 120

QY 121 YARDLGTWTVKTSAGHVATAKYLILATGLLHRKHHTPALPGLADFNKGKVIHSSAWHEDFDA 180
|||||
Db 121 YARDLGTWTVKTSAGHVATAKYLILATGLLHRKHHTPALPGLADFNKGKVIHSSAWHEDFDA 180

QY 181 EGQRVAVIGAGATSIQIVQELAKKADQVTMFMRPPSYCLPMRQRTMDRNEQTAWKAYYPT 240
|||||
Db 181 EGQRVAVIGAGATSIQIVQELAKKADQVTMFMRPPSYCLPMRQRTMDRNEQTAWKAYYPT 240

QY 241 LFEASRKSRIGFPVQAPSVGIFEVSPEQREAYFEELWERGAFNFLACQYREVMVDKKANR 300
|||||
Db 241 LFEASRKSRIGFPVQAPSVGIFEVSPEQREAYFEELWERGAFNFLACQYREVMVDKKANR 300

QY 301 LVYDFWAKKTRSRIVNPAKRDLMAPLEPPYWFGTKRSPLESDDYYEMLDKPSVEIVNLEQS 360
|||||
Db 301 LVYDFWAKKTRSRIVNPAKRDLMAPLEPPYWFGTKRSPLESDDYYEMLDKPSVEIVNLEQS 360

QY 361 PIVAVTKTGVLSDGSKRECDTIVLATGFDSDFTGSLTHMGLKNKKGVDLKEVWKDGISTY 420
|||||
Db 361 PIVAVTKTGVLSDGSKRECDTIVLATGFDSDFTGSLTHMGLKNKKGVDLKEVWKDGISTY 420

QY 421 MGVSFHGFPNAFFVATAQAAPTIVLSNGPTIIETQVDLIADTIAKLEAEHATSVEATKSAQE 480
|||||
Db 421 MGVSFHGFPNAFFVATAQAAPTIVLSNGPTIIETQVDLIADTIAKLEAEHATSVEATKSAQE 480

QY 481 AWSIMIAKMNEHTLFPLTDSWWTGGNIPGKATRALTFIGGIALYEIQCOEKVANWDGFDV 540
|||||
Db 481 AWSIMIAKMNEHTLFPLTDSWWTGGNIPGKATRALTFIGGIALYEIQCOEKVANWDGFDV 540
```

QY 541 LHAPC 545
|||||
Db 541 LHAPC 545

RESULT 8
US-09-677-682B-3
; Sequence 3, Application US/09677682B
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; TITLE OF INVENTION: Detoxification
; FILE REFERENCE: 35718/204101
; CURRENT APPLICATION NUMBER: US/09/677,682B
; CURRENT FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 545
; TYPE: PRT
; ORGANISM: Exophiala spinifera
US-09-677-682B-3

Query Match 100.0%; Score 2902; DB 20; Length 545;
Best Local Similarity 100.0%; Pred. No. 1.5e-300;
Matches 545; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSATSNRSGDCSVACDAIIVGAGLSGISAAYKLRKLRLNAKIFEGAPDFGGVWHNRYPG 60
|||||
Db 1 MSATSNRSGDCSVACDAIIVGAGLSGISAAYKLRKLRLNAKIFEGAPDFGGVWHNRYPG 60

QY 61 ARVDSETPFYQLNIPEVWKDWTWSCRYPDQKELLSYVHHCDKIRGLRKDVYFGAEVVDAR 120
|||||
Db 61 ARVDSETPFYQLNIPEVWKDWTWSCRYPDQKELLSYVHHCDKIRGLRKDVYFGAEVVDAR 120

QY 121 YARDLGTWTVKTSAGHVATAKYLILATGLLHRKHTPALPLGLADFNKGVIHSSAWHEDFDA 180
|||||
Db 121 YARDLGTWTVKTSAGHVATAKYLILATGLLHRKHTPALPLGLADFNKGVIHSSAWHEDFDA 180

QY 181 EGQRVAVIGAGATSIQIVQELAKKADQVTFMRRPSYCLPMRQRTMDRNEQTAWKAYYPT 240
|||||
Db 181 EGQRVAVIGAGATSIQIVQELAKKADQVTFMRRPSYCLPMRQRTMDRNEQTAWKAYYPT 240

QY 241 LFEASRKSRIIGFPVQAPSVGIFEVSPEQREAYFEELWERGAFNFLACQYREVMDKKANR 300
|||||
Db 241 LFEASRKSRIIGFPVQAPSVGIFEVSPEQREAYFEELWERGAFNFLACQYREVMDKKANR 300

QY 301 LVYDFWAKKTRSRIVNPAKRDLMALEPPYWFGTKRSPLESYYEMLDKPSVEIVNLEQS 360
|||||
Db 301 LVYDFWAKKTRSRIVNPAKRDLMALEPPYWFGTKRSPLESYYEMLDKPSVEIVNLEQS 360

QY 361 PIVAVTKTGVLLSDGSKRECDTIVLATGDFSFTGSLTHMGLKNKHGVDLKEVWKDGISTY 420
|||||
Db 361 PIVAVTKTGVLLSDGSKRECDTIVLATGDFSFTGSLTHMGLKNKHGVDLKEVWKDGISTY 420

QY 421 MGVSFHGFPNAPFFVATAQAAPTIVLSNGPTIETQVDLIADTIAKLEAEHATSVKSAQE 480
|||||
Db 421 MGVSFHGFPNAPFFVATAQAAPTIVLSNGPTIETQVDLIADTIAKLEAEHATSVKSAQE 480

QY 481 AWSIMIAKMNHEHTLFPLTDSWWTGGNIPGKATRALTFIGGIALYEQICQEKVANWDGFDV 540
|||||
Db 481 AWSIMIAKMNHEHTLFPLTDSWWTGGNIPGKATRALTFIGGIALYEQICQEKVANWDGFDV 540

QY 541 LHAPC 545
|||||

Db 541 LHAPC 545

RESULT 9
US-09-882-694-3
; Sequence 3, Application US/09882694
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; TITLE OF INVENTION: Detoxification
; FILE REFERENCE: 35718/208255
; CURRENT APPLICATION NUMBER: US/09/882,694
; CURRENT FILING DATE: 2001-06-12
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 545
; TYPE: PRT
; ORGANISM: Exophiala spinifera
US-09-882-694-3

Query Match 100.0%; Score 2902; DB 22; Length 545;
Best Local Similarity 100.0%; Pred. No. 1.5e-300;
Matches 545; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSATSNRSGDCSVACDAIIVGAGLSGISAAYKLRKLRLNAKIFEGAPDFGGVWHNRYPG 60
|||||
Db 1 MSATSNRSGDCSVACDAIIVGAGLSGISAAYKLRKLRLNAKIFEGAPDFGGVWHNRYPG 60

QY 61 ARVDSETPFYQLNIPEVWKDWTWSCRYPDQKELLSYVHHCDKIRGLRKDVYFGAEVVDAR 120
|||||
Db 61 ARVDSETPFYQLNIPEVWKDWTWSCRYPDQKELLSYVHHCDKIRGLRKDVYFGAEVVDAR 120

QY 121 YARDLGTWTVKTSAGHVATAKYLILATGLLHRKHTPALPLGLADFNKGVIHSSAWHEDFDA 180
|||||
Db 121 YARDLGTWTVKTSAGHVATAKYLILATGLLHRKHTPALPLGLADFNKGVIHSSAWHEDFDA 180

QY 181 EGQRVAVIGAGATSIQIVQELAKKADQVTFMRRPSYCLPMRQRTMDRNEQTAWKAYYPT 240
|||||
Db 181 EGQRVAVIGAGATSIQIVQELAKKADQVTFMRRPSYCLPMRQRTMDRNEQTAWKAYYPT 240

QY 241 LFEASRKSRIIGFPVQAPSVGIFEVSPEQREAYFEELWERGAFNFLACQYREVMDKKANR 300
|||||
Db 241 LFEASRKSRIIGFPVQAPSVGIFEVSPEQREAYFEELWERGAFNFLACQYREVMDKKANR 300

QY 301 LVYDFWAKKTRSRIVNPAKRDLMALEPPYWFGTKRSPLESYYEMLDKPSVEIVNLEQS 360
|||||
Db 301 LVYDFWAKKTRSRIVNPAKRDLMALEPPYWFGTKRSPLESYYEMLDKPSVEIVNLEQS 360

QY 361 PIVAVTKTGVLLSDGSKRECDTIVLATGDFSFTGSLTHMGLKNKHGVDLKEVWKDGISTY 420
|||||
Db 361 PIVAVTKTGVLLSDGSKRECDTIVLATGDFSFTGSLTHMGLKNKHGVDLKEVWKDGISTY 420

QY 421 MGVSFHGFPNAPFFVATAQAAPTIVLSNGPTIETQVDLIADTIAKLEAEHATSVKSAQE 480
|||||
Db 421 MGVSFHGFPNAPFFVATAQAAPTIVLSNGPTIETQVDLIADTIAKLEAEHATSVKSAQE 480

QY 481 AWSIMIAKMNHEHTLFPLTDSWWTGGNIPGKATRALTFIGGIALYEQICQEKVANWDGFDV 540
|||||
Db 481 AWSIMIAKMNHEHTLFPLTDSWWTGGNIPGKATRALTFIGGIALYEQICQEKVANWDGFDV 540

QY 541 LHAPC 545
|||||
Db 541 LHAPC 545

```

RESULT 10
US-09-882-694A-3
; Sequence 3, Application US/09882694A
; GENERAL INFORMATION:
; APPLICANT: Duwick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; TITLE OF INVENTION: Detoxification
; FILE REFERENCE: 5718-111
; CURRENT APPLICATION NUMBER: US/09/882,694A
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 545
; TYPE: PRT
; ORGANISM: Exophiala spinifera
US-09-882-694A-3

Query Match      100.0%; Score 2902; DB 22; Length 545;
Best Local Similarity 100.0%; Pred. No. 1.5e-300;
Matches 545; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSATSNSRGDCSVACDAIIVGAGLSGISAVYKLRKRLNAKIFEGAPDFGGVWHNRYPG 60
Db 1 MSATSNSRGDCSVACDAIIVGAGLSGISAVYKLRKRLNAKIFEGAPDFGGVWHNRYPG 60

QY 61 ARVDSETPFYQLNIPEVWKDWTWSCRYPDQKELLSYVHCHDKIRGLRKDVYFGAEVVDAR 120
Db 61 ARVDSETPFYQLNIPEVWKDWTWSCRYPDQKELLSYVHCHDKIRGLRKDVYFGAEVVDAR 120

QY 121 YARDLGTWTVKTSAGHVATAKYLIATGLLHRKHHTPALPGLADFNKGKVIHSSAWHEDFDA 180
Db 121 YARDLGTWTVKTSAGHVATAKYLIATGLLHRKHHTPALPGLADFNKGKVIHSSAWHEDFDA 180

QY 181 EGQRVAVIGAGATSIQIVQELAKKADQVTMFMRRPSYCLPMRQRTMDRNEQTAWKAYYPT 240
Db 181 EGQRVAVIGAGATSIQIVQELAKKADQVTMFMRRPSYCLPMRQRTMDRNEQTAWKAYYPT 240

QY 241 LFEASRKSRIIGFPVQAPSVGIFEVSPEQREAYFEELWERGAFNFLACQYREVMVDKKNR 300
Db 241 LFEASRKSRIIGFPVQAPSVGIFEVSPEQREAYFEELWERGAFNFLACQYREVMVDKKNR 300

QY 301 LVYDFWAKKTRSRIVNPAKRDLMAPLEPPYWFGTKRSPLESDDYEMLDKPSVEIVNLEQS 360
Db 301 LVYDFWAKKTRSRIVNPAKRDLMAPLEPPYWFGTKRSPLESDDYEMLDKPSVEIVNLEQS 360

QY 361 PIVAVTKTGVLSDGSKRECDTIVLATGFDSTGSLTHMGLKNKHGVLDLKEVWKDGISTY 420
Db 361 PIVAVTKTGVLSDGSKRECDTIVLATGFDSTGSLTHMGLKNKHGVLDLKEVWKDGISTY 420

QY 421 MGVSFHGFPNFAFFVATAQAAPTIVLSNGPTIETQVDLIADTIKLEAEHATSVEATKSAQE 480
Db 421 MGVSFHGFPNFAFFVATAQAAPTIVLSNGPTIETQVDLIADTIKLEAEHATSVEATKSAQE 480

QY 481 AWSIMIAKMNEHTLFPLTDSWWTGGNIPGKATRALTFIGGIALYEQICQEKVANWDGFDV 540
Db 481 AWSIMIAKMNEHTLFPLTDSWWTGGNIPGKATRALTFIGGIALYEQICQEKVANWDGFDV 540

QY 541 LHAPC 545
Db 541 LHAPC 545

RESULT 11
US-09-504-358-22
; Sequence 22, Application US/09504358
```

```

; GENERAL INFORMATION:
; APPLICANT: Rouviere, Pierre E.
; APPLICANT: Brzostowicz, Patricia C.
; TITLE OF INVENTION: GENES AND ENZYMES FOR THE PRODUCTION OF ADIPIC ACID INTERMED.
; FILE REFERENCE: BC1001 US NA
; CURRENT APPLICATION NUMBER: US/09/504,358
; CURRENT FILING DATE: 2000-02-15
; EARLIER APPLICATION NUMBER: 60/120,702
; EARLIER FILING DATE: 1999-February-19
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 22
; LENGTH: 529
; TYPE: PRT
; ORGANISM: Brevibacterium sp HCU
US-09-504-358-22

Query Match      28.7%; Score 833.5; DB 19; Length 529;
Best Local Similarity 32.5%; Pred. No. 4.2e-79;
Matches 167; Conservative 107; Mismatches 237; Indels 3; Gaps 3;

QY 5 SNSRGDCSVACDAIIVGAGLSGISAVYKLRKRLNAKIFEGAPDFGGVWHNRYPGARVD 64
Db 11 AQANADETEVLDAIIVGGFSGPVSDRLREDGFKYKVWDAAGGFGGIWWNCYPGARTD 70

QY 65 SETPFYQLNIPEVWKDWTWSCRYPDQKELLSYVHCHDKIRGLRKDVYFGAEVVDARYARD 124
Db 71 STGQIYQFOYKDLWKDFDKELYPDFENGVEYFEYVDSQLDSRDVTFTFAESCTWDDA 130

QY 125 LGTWTVTKTSAGHVATAKYLIATGLLHRKHHTPALPGLADFNKGKVIHSSAWHE-DFDAEQ 183
Db 131 AKWTVRSSEGREQARAVIVATGFGAKPLYPNIEGLDSFEGECHHTARWPQGGLDMTGK 190

QY 184 RVAVIGAGATSIQIVQELAKKADQVTMFMRRPSYCLPMRQRTMDRNEQTAWKAYYPTLFE 243
Db 191 RVVVMGTGASGIQVIOEAAVAEHLTVFQRTPNLALPMRQRLSADDNDRYRENIEDRFQ 250

QY 244 ASRKSRIIGFPVQAPSVGIFEVSPEQREAYFEELWERGAFNFLACQYREVMVDKKNRLVY 303
Db 251 IRDNSFAGDFYFIPONAAADTPEDERTAIYEKMWDEGGFPLWLGNFQGLLTDEAANHTFY 310

QY 304 DFWAKKTRSRIVNPAKRDLMAPLEPPYWFGTKRSPLESDDYEMLDKPSVEIVNLEQSPIV 363
Db 311 NFWRSKVHDRVKDPKTAEMLAPATPPHPFGVKRPSLEQNYFDVYNQDNVDLIDSNATPIT 370

QY 364 AVTKTGVLSDGSKRECDTIVLATGFDSTGSLTHMGLKNKHGVLDLKEVWKDGISTYMGV 423
Db 371 RVLPNGVETPDGVV-ECDVLVLATGFDNNSGGINAIDIK-AGQLLRDKWATGVDITYMGL 428

QY 424 FSHGFPNFAFFVATAQAAPTIVLSNGPTIETQVDLIADTIKLEAEHATSVEATKSAQEAWS 483
Db 429 STHGFPNLMFLYGPQSPSGFCNGTDFGGAPGCDMWADVFLWLKDNKGISRFESTEEVERWR 488

QY 484 IMIAKMNEHTLFPLTDSWWTGGNIPGKATRALTF 517
Db 489 AHVDDIFVNSLFPKAKSWYWGAVPGKPAQMLNY 522

RESULT 12
US-09-954-314-22
; Sequence 22, Application US/09954314
; GENERAL INFORMATION:
; APPLICANT: Rouviere, Pierre E.
; APPLICANT: Brzostowicz, Patricia C.
; TITLE OF INVENTION: GENES AND ENZYMES FOR THE PRODUCTION OF ADIPIC ACID INTERMED.
; FILE REFERENCE: BC1001 US NA
; CURRENT APPLICATION NUMBER: US/09/954,314
; CURRENT FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: 60/120,702
; PRIOR FILING DATE: 1999-February-19
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: Microsoft Office 97
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2002, 05:55:03 ; Search time 43.81 Seconds
(without alignments)
303.856 Million cell updates/sec

Title: US-09-882-694A-3
Perfect score: 2902
Sequence: 1 MSATNSNRGDCSVACDAIIV.....QICQEKVANWDGFDVYLHAPC 545

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	202.5	7.0	520	4 US-09-030-995-3	Sequence 3, Appli
2	171.5	5.9	532	2 US-08-676-841-8	Sequence 8, Appli
3	170.5	5.9	532	2 US-08-560-916-8	Sequence 8, Appli
4	161.5	5.6	95	4 US-09-655-270A-26	Sequence 26, Appli
5	123.5	4.3	88	4 US-09-655-270A-29	Sequence 29, Appli
6	120.5	4.2	455	4 US-09-030-995-2	Sequence 2, Appli
7	118	4.1	92	4 US-09-655-270A-28	Sequence 28, Appli
8	116	4.0	73	4 US-09-655-270A-32	Sequence 32, Appli
9	100.5	3.5	491	4 US-09-262-856A-5	Sequence 5, Appli
10	99.5	3.4	544	1 US-08-472-028A-8	Sequence 8, Appli
11	99.5	3.4	544	2 US-08-808-931-8	Sequence 8, Appli
12	99.5	3.4	544	3 US-08-808-323-8	Sequence 8, Appli
13	99.5	3.4	544	3 US-09-050-603A-8	Sequence 8, Appli
14	99.5	3.4	544	3 US-09-102-420B-8	Sequence 8, Appli
15	99.5	3.4	544	4 US-09-071-296-8	Sequence 8, Appli
16	99.5	3.4	544	4 US-09-196-268-8	Sequence 8, Appli
17	99.5	3.4	544	4 US-09-015-683-8	Sequence 8, Appli
18	99.5	3.4	544	4 US-09-191-998-8	Sequence 8, Appli
19	99.5	3.4	544	4 US-09-497-698-8	Sequence 8, Appli
20	97.5	3.4	51	4 US-09-655-270A-27	Sequence 27, Appli
21	97	3.3	52	4 US-09-655-270A-34	Sequence 34, Appli
22	95.5	3.3	382	1 US-07-783-705A-3	Sequence 3, Appli
23	95	3.3	615	4 US-09-297-937C-9	Sequence 9, Appli
24	95	3.3	1276	4 US-09-297-937C-13	Sequence 13, Appli
25	93	3.2	711	1 US-08-235-838-7	Sequence 7, Appli
26	93	3.2	711	2 US-08-465-473B-7	Sequence 7, Appli
27	92.5	3.2	551	3 US-08-886-886-15	Sequence 15, Appli

28	92	3.2	549	3	US-08-886-886-13	Sequence 13, Appli
29	91.5	3.2	921	4	US-09-206-800-11	Sequence 11, Appli
30	91.5	3.2	1150	2	US-08-589-756-3	Sequence 3, Appli
31	91.5	3.2	1150	4	US-09-206-800-3	Sequence 3, Appli
32	91.5	3.2	1150	4	US-09-206-898-3	Sequence 3, Appli
33	91	3.1	378	4	US-09-199-637A-367	Sequence 367, App
34	89	3.1	1007	2	US-08-551-459-4	Sequence 4, Appli
35	89	3.1	4150	4	US-09-428-517-2	Sequence 2, Appli
36	86.5	3.0	1181	4	US-09-206-898-23	Sequence 23, Appli
37	86	3.0	364	3	US-08-817-926-51	Sequence 51, Appli
38	86	3.0	364	4	US-09-203-893A-2	Sequence 2, Appli
39	86	3.0	364	4	US-09-203-893A-4	Sequence 4, Appli
40	86	3.0	364	4	US-09-203-893A-6	Sequence 6, Appli
41	86	3.0	364	4	US-09-203-893A-8	Sequence 8, Appli
42	86	3.0	364	4	US-09-203-893A-10	Sequence 10, Appli
43	86	3.0	364	4	US-09-203-893A-12	Sequence 12, Appli
44	86	3.0	364	4	US-09-203-893A-14	Sequence 14, Appli
45	86	3.0	364	4	US-09-203-893A-16	Sequence 16, Appli

ALIGNMENTS

RESULT 1
US-09-030-995-3
; Sequence 3, Application US/09030995
; Patent No. 6235519
; GENERAL INFORMATION:
; APPLICANT: Wang, Yongzhao
; APPLICANT: Childs, John D.
; APPLICANT: Squires, Charles H.
; TITLE OF INVENTION: A Gene Involved in Thiophene
; FILE REFERENCE: EBC97-05
; CURRENT APPLICATION NUMBER: US/09/030,995
; CURRENT FILING DATE: 1998-02-26
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 520
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-030-995-3

Query Match	7.0%	Score 202.5;	DB 4;	Length 520;
Best Local Similarity	21.9%;	Pred. No. 1.4e-12;		
Matches 108;	Conservative 81;	Mismatches 194;	Indels 111;	Gaps 22;
QY	19	IVGAGLSGISAVYKLRKLRLNAKIFEGAPDFGGVHWNRYPGARVDSETPFYQLNIPEVW	78	
Db	7	VIGGGVSLSSIKCCVEEGLEPVCFTERTDDIGLWRFQENP---EEGRASIYKSVIINTS	63	
QY	79	KDWTWSCRYPDQKELLSVHHCDKIRGLR-----KDYFGEAEVVDARYARDL---G	126	
Db	64	KEMMCFSDYPIPDHYPNFMHNAQVLEYFRMYAKEFDLLKYIRFKTTVCSVKKQPDFATSG	123	
QY	127	TWTVKT-SAGHVATAKY--LILATGLLHRKHTP--ALPGLADFNKGVHSSAWHEDFDAE	181	
Db	124	QWEVVTSESEKKEMNVFDGMVCTGHTNAHLPLESFPFGIEKPKQYFHSRDYKNPEGFT	183	
QY	182	QORVAVIGAGATSIQIVQELAKKADQVTMFMRRPSYCLPMRQRTMDRNEQTAWKAYPTL	241	
Db	184	GKRVIIGIGNSGDLAVEISQTAKQVFLSTRRGAWIL-----	221	
QY	242	FEASRKSRIQFPVQAPSVGIFEVSPEQREAYFEELWERGAFNFLACQYREVMVDKKAN-R	300	
Db	222	---NRVGDYGYPA-----DVLFSRLTHF--IW-KICGQSLANKY----LEKKINQR	263	
QY	301	LVYDFWAKKTRSRIVNPAKRLDMLAPLEPPYWFGTGKRSPLSDYYEMLDKPSVEIVNL--E	358	
Db	264	FDHEMFGKPKHRAIS-----QHTFLND-----DLPNRIISGLVKV	299	

Db 123 ENVV-----RLVTWDENTRKFTVTVDLPNDHCYSEDFDNVIVASGHFSTPNVPEEPG 175
QY 161 LADFNKGVIHSSAWHEDEDAEGORVAVIGAGATSIQVQELAKKADOVTFMFMRRPSYCLP 220
Db 176 FDQFNGRILHAHDFRDAKEFIGKDVLLIGT-----SY--- 207
QY 221 MRQRTMDRNEQTANKAYPYPTLFEASRKSRI GPPVQAPSVGIFEVSPEQREAYFEELWERG 280
Db 208 -----SAEDIG-----SQCWKYG 220
QY 281 AFNFLACQYREVMVDKKNRLVYDFWAKKTRSRIVNPAKRDLMAPLEPPYWFGTKRSPLE 340
Db 221 ANSITNC-YR-----TKPMGY-----HW----- 237
QY 341 SDYYEMLDKPVSVEIVNLEQSPIVAVTKTGVLSSDGSKRECDTIVLATGF-DSFTGSLTHM 399
Db 238 PDNWE--EKPLLEKVVHNTA-----TFKDGSTKQIDAIIILCTGYKHHFPEFLPDDL 285
QY 400 GLKNKHGVDLKEVWKDGIS-----TYMGVFSHGFPNAPFFVATA-QAPTVLSNGPTIIE 451
Db 286 RLRTANRLATADLYK-GVAYVHNPALEFYIGMQDQWFTFNMFDAQAQAWWARDVILGRIALPA 344
QY 452 TQVDLIADTIAKLEAEHA 469
Db 345 GKQEMVADVEARVAEDA 362

RESULT 7
US-09-655-270A-28
; Sequence 28, Application US/09655270A
; Patent No. 6329151
; GENERAL INFORMATION:
; APPLICANT: Rouviere, Pierre E.
; TITLE OF INVENTION: High Density Sampling of Differentially Expressed Prokaryotic mR
; FILE REFERENCE: BC1011 US NA
; CURRENT APPLICATION NUMBER: US/09/655,270A
; CURRENT FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: 60/120,702
; PRIOR FILING DATE: 1999-February-19
; PRIOR APPLICATION NUMBER: 60/152,542
; PRIOR FILING DATE: 1999-September-03
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 28
; LENGTH: 92
; TYPE: PRT
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: microbial enrichment culture-
; OTHER INFORMATION: not one single organism
US-09-655-270A-28

Query Match 4.1%; Score 118; DB 4; Length 92;
Best Local Similarity 32.6%; Pred. No. 7.1e-05;
Matches 30; Conservative 17; Mismatches 43; Indels 2; Gaps 2;
QY 450 IETQVDLIADTIAKLEAEHATSVEATKSAQEAWSIMIAKMNEHTLFPLTDSWWTGGNIPG 509
Db 1 IETQVEWISDTVPTPSATRSVRSNPPRS-RGGVDADLHRHREPTLFTRGDSWIFGANVPG 59
QY 510 KATRALTFIGGIALYEICQEKVA-NWDGFDV 540
Db 60 KKPSVLFYLGGLGNRYNVLGAVVADSYRGFEL 91

RESULT 8
US-09-655-270A-32
; Sequence 32, Application US/09655270A
; Patent No. 6329151
; GENERAL INFORMATION:
; APPLICANT: Rouviere, Pierre E.

; TITLE OF INVENTION: High Density Sampling of Differentially Expressed Prokaryoti
; FILE REFERENCE: BC1011 US NA
; CURRENT APPLICATION NUMBER: US/09/655,270A
; CURRENT FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: 60/120,702
; PRIOR FILING DATE: 1999-February-19
; PRIOR APPLICATION NUMBER: 60/152,542
; PRIOR FILING DATE: 1999-September-03
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 32
; LENGTH: 73
; TYPE: PRT
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: microbial enrichment cultur
; OTHER INFORMATION: not one single organism
US-09-655-270A-32

Query Match 4.0%; Score 116; DB 4; Length 73;
Best Local Similarity 41.4%; Pred. No. 7.8e-05;
Matches 29; Conservative 5; Mismatches 36; Indels 0; Gaps 0;
QY 87 YPDQKELLSYVHHCDKIRLKRKDVYFGAEVVDARYARDLGTWTVKTSAGHVATAKYLILA 146
Db 4 YITQPEILEYLEDVVDREDFLRRTFRFGTEVKVSATYLEDEGLWEVTTGGGAVYRAKYVINA 63
QY 147 TGLLHRKHTP 156
Db 64 VGLLSAINEP 73

RESULT 9
US-09-262-856A-5
; Sequence 5, Application US/09262856A
; Patent No. 6333164
; GENERAL INFORMATION:
; APPLICANT: TAKESAKO, Kazutoh
; APPLICANT: MIZUTANI, Shigetoshi
; APPLICANT: ENDO, Masahiro
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: FUNGAL ANTIGENS AND PROCESS FOR PRODUCING THE SAME
; FILE REFERENCE: 1422-372P
; CURRENT APPLICATION NUMBER: US/09/262,856A
; CURRENT FILING DATE: 1999-03-04
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 491
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-262-856A-5

Query Match 3.5%; Score 100.5; DB 4; Length 491;
Best Local Similarity 18.3%; Pred. No. 0.087;
Matches 100; Conservative 76; Mismatches 178; Indels 193; Gaps 21;
QY 16 DAIIVGAGLSGISAVYKLRKLRLNAKIFEGAPDFGGVHWNRYPCARVDSETFFYQLN-- 73
Db 27 DVVVIGGGPGGYVAAIKAAQLGLNTACIEKRGALGGTC-----LNVG 68
QY 74 -IPEVWKDWTWSCRYPDQKELLSYVHHCDKIRGLR-----KDVYF 112
Db 69 CIP-----SKSLNNSHLHLHQIQHEAKERGISIQGEVGVDFPKLMAAKEKAVKQLTG 120
QY 113 GAEVVDAR----YARDLGTW---TVKT-----SAGHVATAKYLILATGLLHRKHTPALP 159
Db 121 GIEMLFKKNKVDYLGAGSFFVNEKTVKVTPIDGSEAQAEVEADHIIVATG---SEPTP-FP 176
QY 160 GLADFNGKVIHSSAWHEDFDAEQRVAVIGAGATSIQIVQELAKKADQVTMFMRRPSYCL 219

Db 177 GI-EIDERIVTSTGILSLKEVPERLALICGGIIGLEMASVYARLGSKVTVI----- 227
QY 220 PMRQRTMDRNEQTAWKAYPTLFEASRKSRIQFPPVQAPSVGIFEVSPQREAYFEELWER 279
Db 228 -----EFQNAIGAGMDAE----- 240
QY 280 GAENFLACQYREVMDKKANRLV----YDFWAKKTRSRIVNPAKRDLMAPLEPPYWFGTK 335
Db 241 -----VAKQSOKLLAKQGLDF---KLGTKVVKGERDGEVVVKIEVEDVKSGK 283
QY 336 RSPLESDYYEML--DKPSVEIVNLEQSPIVAVTKTGVLSSDGSKRECDTIVLATGDFSFT 393
Db 284 KSDLEADVLLVAIGRRPFTEGLNFEALGLEKDNKGRLLIIDDOFKTKHDHI-RVIGDVTFG 342
QY 394 GSLTHMG-----LKNKH-----GVDLKEVWMDGISTYMG 422
Db 343 PMLAHKAEEGIAAAEYIKKGHHVNVANIPSVMYTHPEVAWVGLNEEQLEQGIKYKVG 402
QY 423 VFSHGFPNAFFVATAQAPTIVLSNGPTI-----IETQ----VDLIADTIAKLEAEHATSVE 473
Db 403 KFP-----FIANSRAKTNMDTDGFKFIADAETQTVLGVHIIIGNAGEMIAEAGLALE 455
QY 474 ATKSAQE 480
Db 456 YGASTED 462

RESULT 10
US-08-472-028A-8
; Sequence 8, Application US/08472028A
; Patent No. 5767373
; GENERAL INFORMATION:
; APPLICANT: Ward, Eric R
; APPLICANT: Volrath, Sandra
; TITLE OF INVENTION: Manipulation of Protoporphyrinogen
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/472,028A
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1748/CIP
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 544 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-472-028A-8

Query Match 3.4%; Score 99.5; DB 1; Length 544;
Best Local Similarity 20.1%; Pred. No. 0.13;
Matches 82; Conservative 47; Mismatches 115; Indels 163; Gaps 18;

QY 19 IVGAGLSGISAVYKLRKLRLNAKIFEGAPDFGGVWHNRYPGARVDSSETPFYQLNIPVW 78
Db 50 VVGAGVSGLAAYRLRQSGVNVTVFEAADRAGG-----KIRNSEGGF-----VW 94
QY 79 KDWT-----WSC-----RYPDQKELLSYV----- 97
Db 95 DEGANTMTEGEWEASRLDIDLGLQDKQYPSNSQHKRYIVKDGAPALIPSDPISLMKSSVL 154
QY 98 -----HHCDKIRGLRKDVYFGAEVVDARYARDLGTWT 129
Db 155 STKSKIALFFEPFLYKKANTRNSGKVSSEHLSVSGFCERHFGREVVD--YFVD--PFV 210
QY 130 VKTSAGHVATAKYLIILATGLLHRKHTPALPGLADFNKGKVIHSSAWHEDFDAEGORVAVIG 189
Db 211 AGTSAGD-----PESLSIRHAFPALWNLERKYGSVI-----VG 243
QY 190 AGATSIQIVQELAKKADQVTMFMRPPSYCLPMRQR---TMDRNEQTAWKAY--YPTLFE 243
Db 244 A-----ILSKLAAGD-----PVKTRHDSGKRRNRVVSFHHGGMQSLIN 284
QY 244 ASRKS-----RIGFPVQA-----PSVGIFEVSPQREAYFEELWERGAFNFL---- 285
Db 285 ALHNEVGDDNVKLGTEVLSLACTFDGVPALGRWSISVDSKDSGDKDLASNOTFDAVIMTA 344
QY 286 -ACQYREVMDKKANRLVYDFWAKK---TRSRIVNPAKR-DLMAPLE 327
Db 345 PLSNVRRMKFTKGGAPVVLDPLPKMDYLPPLSLMVTAFKKDDVKKPLE 391

RESULT 11
US-08-808-931-8
; Sequence 8, Application US/08808931
; Patent No. 5939602
; GENERAL INFORMATION:
; APPLICANT: Volrath, Sandra
; APPLICANT: Johnson, Marie
; APPLICANT: Potter, Sharon
; APPLICANT: Ward, Eric
; APPLICANT: Heifetz, Peter
; TITLE OF INVENTION: DNA Molecules Encoding Plant
; TITLE OF INVENTION: Protoporphyrinogen Oxidase and Inhibitor-Resistant Mutants
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5939602artis Corporation
; STREET: 520 White Plains Road, P.O. Box 2005
; CITY: Tarrytown
; STATE: NY
; COUNTRY: USA
; ZIP: 10591-9005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/808,931
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/012,705
; FILING DATE: 28-FEB-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/013,612
; FILING DATE: 28-FEB-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/020,003
; FILING DATE: 21-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: CGC 1847
; TELECOMMUNICATION INFORMATION:

